Gender/Genre: Gender difference in disciplinary communication

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Brian N. Larson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Mary Lay Schuster, adviser

May, 2015

Acknowledgements

So many of those I would like to thank for getting me to this point are my teachers. Some were my mentors a long time ago: Teachers who shaped my character—including Beverly Burmeister and Ron Pethoud—those who helped me through tough times—including Patricia Cox-Charles and Rita Clavadatscher—and those who were friends and guides of one kind or another—including Dr. Susan Brantly and Dr. Andrew Sihler.

Others have had a more recent effect on my work as a scholar, shaping my interests in rhetoric and in disciplinary, professional, and technical communication: Dr. Lee-Ann Kastman Breuch, who introduced me to methods in the abstract and then guided me in their use as her research assistant; Dr. Christina Haas, who challenged me to sharpen my thinking and my research questions; Dr. Serguei Pakhomov, who showed me NLP and helped me interrogate its capabilities; Dr. Carol Berkenkotter, whose warmth and eager guidance opened the world of genre theory to me; and Dr. Anne Lazaraton, whose meticulous methodological care is an inspiration to me.

A special acknowledgement must go to Dr. Mary Schuster, my adviser throughout my Ph.D. program. Indeed, her advice began before I came to my program, with suggestions about the form that my application should take. It would be hard to overstate the value of such a pragmatic mentor with high standards and great compassion. She has always looked for ways that my work could be better, always taking into account the limitations of the 24-hour day.

Other folks in the Writing Studies Department have been important, too. My graduate-school colleague Laura Pigozzi has played an important role in my development as a scholar; our conversations about our work have helped me understand how to make my own efforts relevant to others who are not all wrapped up in them and to treasure the efforts of others. She has also become a close friend. I have always felt that

Director of Graduate Studies Dr. Donald Ross was dedicated to the success of the grad students in his charge; and Graduate Programs Administrator Nan Nelson has been an indispensable resource for solving the thousand administrative and technical problems that crop up in five years of graduate school. I hope I can call *them* my friends, too.

While centering on folks in the Department of Writing Studies, I should pause to express my profound gratitude to the department and the College of Liberal Arts at the University of Minnesota for financial support that I have received throughout my Ph.D. training, and particularly for the James I. Brown Summer Research Fellowship in Rhetoric in 2014 and the Graduate Research Partnership Program grant (supervised by Dr. Breuch) in summer 2012. These additional resources allowed me to provide financial incentives to study participants and to pay a research assistant to help with manual annotation of the text samples.

Outside of school, there are many individuals whom I would like to thank for their support, friendship and love over the years. Judith Lindenau comes to mind first. Though we lost her last year, it's hard to think with sadness about the death of one for whom life remained an adventure for 73 years. Judith helped me do the hard work of evaluating my options before I came to graduate school. I can remember her telling me: "Don't do what you think you *should* want to do; do what you *want* to do." Despite that admonition, Judith had me running around, gathering information about my options, to make sure that whatever decision I made, it would be supported by the data. I love you and miss you, my dear!

My colleagues at Larson Skinner, PLLC, the law firm that I founded 13 years ago, always supported me through this process. Mitch Skinner, Shannon Larson, and Elizabeth Sobotka are due my thanks and affection.

My family has also been very supportive, though from time to time, a little uncertain about what exactly I've been doing. My thanks to my mom and dad for all that they've done for me. My spouse, Robert Tendal, has tolerated so much from me over the last five years—actually, over the last 26 years—largely without complaint. I've dedicated this dissertation to him, and if you knew us, you'd know that I could not have chosen otherwise.

Finally, to the two folks who inspired my interest in research and teaching in the first place: my thanks to Dr. Genelle Belmas, a long-time friend and my first research collaborator, for whetting my appetite for research; and to Professor Brad Clary, who invited me to teach for the first time and helped me realize that I want to do it until I die.

Brian
May 27, 2015
Minneapolis, MN

Dedication

To Bob:

Without you, my possibilities would all have been disappointments. With you, more have become real than I could have dreamed.

—B.N.L.

Abstract

Within the professions, writers are expected to express themselves in certain ways, often within genres that are bound by conventions, including linguistic register. The student entering a profession learns those genres as if they are mandatory and static, and conforming or failing to conform to conventions is believed to have ties to career consequences. However, new members of a profession come to it with other habitual language practices affected—according to previous research—by the writer's gender. Rhetorical genre theory and disciplinary, professional, and technical communication theory do not offer a full account for the ways in which these old habits and new conventions must interact, and previous research in gender and language does not fully account for how gendered persons write when confronted with high-stakes convention-bound writing tasks. I used tools from statistics and natural language processing (NLP) to assess stylistic features that previous research has associated with gender differences in written language: I applied those tools to texts created by law students near the end of their first year of study in the genre of a court memorandum, and I found there was no pattern of difference between male and female writers in these texts.

I propose a "cognitive pragmatic rhetorical" (CPR) theory, grounded in work of Straßheim (2010), who attempted to bridge the relevance philosophy of Alfred Schutz (Schutz, 1964, 1966, 1973) and the Relevance Theory of Sperber and Wilson (1995); I have extended Straßheim's work with insights from rhetoric and cognitive science. CPR theory explains that these apprentice members of a professional community will expend great effort to conform to its conventions and genres because of the students' goals and the practical effects that depend on conformity. Consequently, we expect them to abandon gendered linguistic habits, at least while they are engaged in early training. This dissertation demonstrates a methodologically rigorous gender-difference study; offers evidence for an "anti-essentialist" view of gender differences in communication; and gives insight into the process by which apprentice members of a profession may adjust their communicative processes in response to their training. It demonstrates the utility of CPR theory and NLP tools in scholarly inquiries in rhetoric and disciplinary, professional, and technical communication.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Within the professions, writers are expected to express themselves in certain ways, often within genres believed to be bound by formal conventions, including linguistic register. The student entering a profession may learn those genres as if they are mandatory and static, and she may believe that conforming or failing to conform to conventions has ties to career consequences. However, new members of a profession come to it with other habitual language practices that vary—according to previous research—with the writer's gender. Rhetorical genre theory and disciplinary, professional, and technical communication theory do not fully account for the ways in which these old habits and new conventions must interact in the individual Writer and Reader.[1] And previous research in gender and language does not offer a full account for how gendered persons write when confronted with high-stakes convention-bound writing tasks.

In this dissertation, I propose a "cognitive pragmatic rhetorical" (CPR) theory, grounded in work of Straßheim (2010), who attempted to bridge the relevance philosophy of Alfred Schutz (Schutz, 1964, 1966, 1973) and the Relevance Theory of Sperber and Wilson (1995); I have extended Straßheim's work with insights from rhetoric and

[1] I employ the practice of referring to the hypothetical "Speaker" or "Writer" and "Hearer" or "Reader" with an initial capital, using feminine pronouns for the former (Speaker=she) and masculine pronouns for the latter (Hearer=he). The reader here should regard Speaker and Writer as interchangeable; I will interchange Hearer and Reader as well. Though I do not take up in this dissertation the question of the Speaker and Hearer being one and the same, I recognize that possibility. See Perelman and Olbrechts-Tyteca (1969).

cognitive science. CPR theory takes account of a Writer's goals and beliefs about the world and explains the efforts the Writer employs in finding, discovering, or inventing her communicative performances with the principle of relevance, which holds that she will expend effort in her writing choices that is commensurate with the accessibility and strength of the goals for which she writes and the effect she expects her writing to have on the cognition of the Readers(s). In other words, relevance is the ratio of desired cognitive effects to the cognitive effort expected to be necessary to produce them. Here, I use "ratio" in the sense of a numerical fraction as a metaphor: I think of cognitive effects as the numerator in the relevance fraction and cognitive effort as the denominator of the fraction. Increasing cognitive effects increases the value of the fraction; increasing cognitive effort decreases the value of the fraction.

In the case studied in this dissertation, the Gender F and Gender M[2] authors attended to the linguistic conventions of the legal writing in which they were engaged as first-year law students carefully enough to obscure any habitual variations connected with gender. CPR theory explains that the social context of the writing, particularly the goals that the student participants wished to achieve—approval of their instructors and subsequent career opportunities—may have warranted significant conscious effort to comply with the professional linguistic conventions. It would therefore not be surprising that the differences found in other studies that did not control for participants' communicative objectives—or to put it another way, that did not control for participants' genre knowledge or goals—were not apparent in the writing studied here.

In short, the CPR-theoretic concept of relevance governs the efforts of the Writer to produce the text and explains the ability of novice legal writers to abandon (at least temporarily) habitual linguistic habits. As I shall argue later, relevance helps to explain other phenomena that are of interest in rhetoric and what I refer to as "disciplinary, professional and technical communication" (DP&TC).

The introduction of this term requires a two-paragraph detour. Though I cannot hope here fully to defend my suggestion that the field of "technical communication" should be recast as the field of DP&TC, I can at least define what I mean by the latter. This is particularly important in light of the fact that "disciplinary" here should not be

[2] See the discussion in Section 3.4.2 beginning at page 81 for an account of the gender construct used in the empirical study in this dissertation. See the discussion in Section 4.2.3 regarding the ascription of the "Gender F" and "Gender M" labels to authors of the texts in the study.

confused with the "disciplines" of the discipline of "writing in the disciplines."[3] As I define it, DP&TC is communication relating to "technical" subject matter prepared by or on behalf of experts in the technical subject matter (SMEs). (The hedge "more-or-less" could fairly be placed before each of these criteria.) I propose to define technical subject matter based on the relationship of author's knowledge to some kind of more broadly defined society. If the author is sharing knowledge that most other folks do not have (e.g., regarding a scientific field or methods for assembling a piece of furniture or means of assessing responsibility for an oil spill) that knowledge is sufficiently "technical" for purposes of this definition.[4]

The three species of DP&TC are distinguished from each other by their audiences. The audiences of disciplinary communication are disciplinary peers of the authors or at least distinguished from society in general by their shared disciplinary knowledge or apprenticeship; this includes students writing for professors in their fields, but also scientists writing for each other in academic journals, and lawyers writing memoranda of law to be read by other lawyers or judges. The audiences of professional communication are members of the same organizational group as the authors or persons interacting with the authors in business or professional transactions; this includes middle managers writing reports for their senior managers, business memoranda and proposals, etc. The audiences of technical communication are persons in an asymmetrical knowledge relationship with the authors who are seeking out the communication to proximally mediate an action or belief; this includes the reader of instructions for assembling a piece of IKEA furniture as well as the layperson client of a lawyer receiving legal advice. Of course, these types of communication can overlap: a staff attorney in a company writing a memorandum of advice to another department is engaging both in professional and technical communication.

[3] In writing in the disciplines, "disciplinary" should be taken to refer to academic disciplines. WID might be best understood as the teaching of writing and other communication skills to university students within courses in academic disciplines other than communications, and where the students are also learning the substance of the academic discipline. See Carter (2007) for a discussion.

[4] I am glossing over the concerns about what "technical subject matter" is, legitimately raised by Miller (1979) and others. Note that Miller gave up on defining scientific and technical communication, in a sense, by claiming that all such definitions "leak badly" (p. 614). While eschewing Aristotelian logic, she seemed to believe that defining a discipline meant compartmentalizing it into Aristotelian categories. See Durack (1997) for an expansive definition of "technical," including technologies like sewing, which is consistent with my own.

The empirical study in this dissertation and the application of CPR theory to it have some fairly definite implications, including providing a demonstration of a methodologically rigorous gender-difference study; evidence for an "anti-essentialist" view of gender differences in communication; and insight into the process by which apprentice members of a profession may adjust their communicative processes in response to their training. I suggest in Chapter 7 that CPR theory will lead to broader implications in rhetoric and in DP&TC. Examples discussed there include nuancing rhetorical analysis and understanding the cognitive dynamics of rhetorical resistance.

As a researcher and teacher in DP&TC, I must attend to the decisions that individual human agents make in their communicative performances. But the effectiveness of those performances from the perspective of Speaker or Writer depends upon strategies that work *across* individuals. Consider the lawyer writing a brief for the nine Justices of the Supreme Court and the technical writer writing a manual for a complex piece of scientific equipment. Each of them is an agent making numerous decisions about a text meant to move or educate an audience of many, some of whom the writer perhaps knows, at least by name, but many of whom the writer does not know at all, and never will. DP&T writers must perforce generalize about their audiences, as Aristotle suggested to rhetors in *The Rhetoric* (Aristotle, 2007). So, too, I explore DP&TC from the perspective of one who wishes to generalize, but not as Aristotle did, from anecdotal experiences and a priori reasoning.[5] My commitments are to quasi-foundational (Lazaraton, 2003) or post-positivist paradigms that permit some assessments of validity and cautious generalization.

I do not mean by this approach to suggest the theoretical frameworks that emphasize social activity are not valuable in themselves. A theory grounded in individual cognition, such as the one for which I advocate in this dissertation, might be accused of abstracting away from social constructs; but social theories of knowledge and theories of social knowledge abstract away from individual agents. Theory *is* abstraction. Instead, I argue the individual perspective is valuable and is complementary to and compatible with theoretical models looking at "mid-level" units of analysis, such as activity systems theory (D. R. Russell, 1997). The work of researchers attending to individual cognition

[5] According to Bertrand Russell, "Aristotle maintained that women have fewer teeth than men; although he was twice married, it never occurred to him to verify this statement by examining his wives' mouths" (B. Russell, 1968, p. 7).

can help to expose nuances in the work of those attending to activity systems as objects of study. For example, D. R. Russell (1997, p. 510) describes an activity system as "object-directed," that is, the system largely shares a common motivation. My intuition is that each "subject" in the activity system has a (slightly, at least) different conception of the object to which the system is directed. The researcher attending to the system must impose her own conception of the object to which the system is directed. This interpretive process can lead to valuable knowledge-making. But understanding nuances in the cognition of individual subjects (or "agents," as I'll often refer to them here) is another interpretive process leading to valuable knowledge.

The cognition of new or apprentice members of disciplinary or professional communities has been the object of many studies. I would like to suggest that the boundary between apprentice members and veteran members is quite fuzzy, and that drawing it at the classroom door is a mistake. My anecdotal experience as a lawyer in legal practice and as a teacher in the law-school classroom leads me to see students' and professionals' knowledge of professional conventions lying on a great many continua. Consider these examples: First, some seasoned "transactional lawyers" (whose practice is directed at contracts and business negotiations) of many years' practice were never more familiar with the genres of the courtroom than while they were law students. Second, a law student clerking with a law firm for a summer may become such an expert with a complex agency filing process that a partner at another law firm will ask if she is her firm's "lead associate" on such filings. Finally, an unsystematic review of briefs filed in federal courts showed me that even lawyers of many years' practice as litigators may employ stylistic practices that are disfavored by the great majority of their peers. And yet, this is expert practice.

I don't contend that law students are fully acculturated to any part of the profession of law, but I do contend that we are all agents with varying degrees of acculturation to new disciplines, communities, professions, and tasks at all times. In all of these situations, we have individual goals and motivations about what we want from the environments. We know we can get some of what we want with communicative performances. In any communicative interaction, the Writer has objectives for her communication, more or less strongly felt, beliefs about the world and the Reader, more or less strongly held, and inventional resources in the form of habitual practices, an unconscious understanding of certain human communication techniques that operate more-or-less automatically in Writer and Reader, and possibly others. If she is a new or apprentice member of a professional community, her beliefs about Readers may be less accurate, on the whole, than one more experienced in the community, of course. The Reader seeks to interpret the Writer's efforts for his own reasons, which probably do not perfectly align with the Writer's. Each of these agents is making meaning with or from the communicative performance of the Writer, in a sense, but it's not the same meaning for both of them.

An apprentice member of a professional community struggles to understand the substance-knowledge of the field but also its conventional communicative performances; she struggles to build genre knowledge (Berkenkotter & Huckin, 1994). Though I anticipate that she is strongly motivated by economic and social pressures to identify and conform to genre conventions, including linguistic register, I suspect her genre knowledge must compete with her habitual communicative practices. Some such practices are probably easily shed because they are consciously performed outside the professional context. For example, students appear ready in the law-school classroom to abandon "text speech"—like "when r u going 2b here?"—that they use in informal contexts. But some pervasive habitual communicative practices evade the Writer's attention; I have helped students note and address speech habits (like "up-talk"—the rising intonation at the end of assertions that makes them sound like questions) that diminish the effectiveness of their communications.

Perhaps the most pervasive set of habitual communicative practices of most agents relates to gender. For good or bad, we adopt identities related to our sex, which we signal using gendered communicative performances. As we engage in these habitual practices every day from a young age, one could imagine that they are both very unconsciously performed and very hard to overcome. It is this line of reasoning that undergirds theories of gender and communication that place men and women in different communicative "cultures" (Maltz & Borker, 1982).

This gives rise to the question—and the empirical study in this dissertation—of how the habitual communicative practices of gender interact with the struggle to acculturate to genre conventions of a new profession. In other words, do gendered writers overcome their gendered communicative habits when writing in genres that are new to them, or do their habits show through? How can we explain the answer?

At a national workshop on rhetorical genre theory in 2013, I briefly described the empirical study in this dissertation—which examines legal memoranda written by law students in response to a hypothetical case problem—to a researcher with an international reputation in genre theory. First, she reacted by saying, "That's just student writing," and I came to understand from her that she did not regard such texts as having the status of a genre. Based on the intuitions I described above, I disagreed with her. With regard to my empirical research questions—whether there would be stylistic differences between texts written by men and those written by women in my study—she responded (with no apparent irony), "There won't be any gender difference because of the conventions of the text genre." She acknowledged that her view was an intuition; and on that intuition she and I agreed.

The findings of the study that I conducted were also consistent with those intuitions. There were no patterns of difference in lexical and quasi-syntactic choices between the authors I have categorized as Gender F and Gender M. But I also want to offer an explanation for why and how stylistic differences that seemed so common in earlier studies appear to have been erased when law students wrote these high-stakes writing assignments with only apprentice knowledge of the genre conventions. I believe that providing an explanation for the journey of these students from writing *gendered* texts to writing *genred* texts can help to explain the cognitive mechanisms for the stability and dynamism of genres (Berkenkotter & Huckin, 1994) and the relation of genre and gender as communicative practices. I did not believe that any available theory in rhetoric or DP&TC could explain these phenomena to my satisfaction, so I developed my own theory, which I call "cognitive pragmatic rhetorical (CPR) theory." And so the motivating question for this dissertation and the empirical study in it is this:

> *How can cognitive pragmatic rhetorical (CPR) theory contribute to our understanding of rhetorical and disciplinary, professional, and technical communication theory and in particular to our theories of gender and genre performances?*

In the following paragraphs, I'll preview the contents of this dissertation and its address of CPR theory in detail. But I can summarize the argument here in brief:

CPR theory takes account of a Writer's goals and beliefs about the world and about the Reader(s); it explains the efforts the Writer employs in finding, discovering, or inventing her communicative performances with the principle of relevance, which holds that she will expend effort in her writing choices that is commensurate with the accessibility and strength of the goals for which she writes and the effect she expects her writing to have on the cognition of the Readers(s). In other words, relevance is the ratio of desired cognitive effects to the cognitive effort expected to be necessary to produce them. In the case studied in this dissertation, the Gender F and Gender M authors attended to the linguistic conventions of the legal writing in which they were engaged as first-year law students carefully enough to obscure any habitual variations previously connected with gender. CPR theory explains that the cognitive context of the writing, particularly the social goals that the student participants may have wished to achieve—approval of their instructors and subsequent career opportunities—warranted significant conscious effort to comply with the professional linguistic conventions. It is therefore not surprising that the differences found in other studies that did not control for participants' communicative objectives—or to put it another way, that did not control for participants' genre knowledge—were not apparent in the writing studied here.

1.2 Preview of following chapters

The balance of this chapter summarizes the chapters that follow. This dissertation explores the research question first by proffering an exposition of CPR theory, which attends to the cognition of individual agents in social situations (Chapter 2); it then summarizes some of the literature that has explored gender differences in language (Chapter 3); it describes in some detail the study design (Chapter 4) and findings (Chapter 5) for the empirical study suggested by the question; and finally, it offers a possible explanation of the findings using CPR theory (Chapter 6) before considering some implications, limitations, and future research questions (Chapter 7).

Chapter 2: CPR theory

I have devised cognitive pragmatic rhetorical (CPR) theory as a basis for describing the written performance of the participants in the empirical study in this dissertation;

it provides a means to explain whether and how the participants here could abandon habitual, gendered linguistic performances when they have presently accessible and strongly held goals tied to conformity with a professional text genre. Chapter 2 describes the antecedents of CPR theory, including classical pragmatics and relevance-theoretic pragmatics. The chapter then explains CPR theory, showing how it takes account of a Speaker's (or Writer's) accessible goals and assumptions when explaining how she selects an utterance to perform.

Chapter 3: Gender difference studies

At the heart of this dissertation is a study of gender-differences in the language of students in a professional-training context. Chapter 3 considers several preparatory matters, including questions about whether we should do such studies; a description of the studies that inspired this empirical project and their limitations; descriptions of several other studies, all of which suffered from similar limitations; and proposals for operationalizing the variables of *gender* and *genre* in the present study. It also explains a presumption underlying my study: that is, that women and men do write differently "in the wild"—that is, in informal contexts.

In Section 3.3, I describe two essays deriving from the same study (Argamon, Koppel, Fine, & Shimoni, 2003; Koppel, Argamon, & Shimoni, 2002), which I frequently refer to together as the Argamon/Koppel 02/03 study in this dissertation. The Argamon/Koppel 02/03 study began with a collection of texts from the British National Corpus (BNC), including newspaper and magazine essays, novels, and other works of fiction and non-fiction. The researchers assessed some stylistic characteristics of these texts. Koppel et al. (2002) reported findings of a machine-learning study on this corpus, which showed that a machine-learning algorithm trained on data like those in this corpus could correctly predict the gender of a previously unseen text from the corpus between 77.3% and 82.6% of the time. (The text at p. 64 and following provides an overview of machine-learning algorithms.) The second essay arising from the Argamon/Koppel 02/03 study, Argamon, Koppel, Fine, and Shimoni (2003) offered a statistical analysis of the same stylistic features in the same corpus, showing that there were significant differences between male an female writers. In both essays, the researchers offered the "involved"/"informational" dimension of Biber (1995) as a framework for understanding the statistical differences. The involved end of the dimension, associated with women by these researchers, displays "interaction between the speaker/writer and the listener/reader, such as first and second person pronouns"; while the informational end, associated with men, exhibits a larger numbers of specifiers and particular types of prepositional phrases (Argamon, Koppel, Fine, & Shimoni, 2003, p. 332).

This dissertation explores the interaction of gender performances and genre performances with participants who are gendered persons responding to a high-stakes writing assignment in a professional-training context after receiving training for an academic year in legal writing. This chapter provides background on previous studies exploring related questions, describes methodological limitations that I hope to avoid with this dissertation, and explains the presumption of difference with which we may safely begin before performing the present study.

Chapter 4: Study design

Given the presumption of gender difference in everyday language that I argued for in Chapter 3, I sought to design a study that would address methodological concerns associated with previous studies. Chapter 4 represents my effort to gather a sufficient number of texts in the same genre, written by individual authors who self-identify for their genders. It shows that I used methods similar to those used in previous studies that found gender difference. And it demonstrates that I did so ethically.

The overarching research question in this dissertation was expressed above:

> *How can cognitive pragmatic rhetorical (CPR) theory contribute to our understanding of rhetorical and technical and professional communication theory and in particular to our theories of gender and genre performances?*

Chapter 4 describes four subsidiary empirical research questions and the means by which this study sought to answer them:

1. Do Gender F and Gender M writers in a disciplinary genre in which they are being trained use lexical and quasi-syntactic stylistic features with relative frequencies that vary in relation to their genders?

2. If so, do the differences appear in interpretable patterns?

3. Can machine-learning algorithms categorize the same texts by author gender based on the same features?

4. If so, do they provide interpretable models?

Chapter 4 shows that I have constructed a study that will permit me to answer these questions.

I gathered data at two midwestern law schools—with the pseudonyms "Academy School of Law" and "Lyceum Law College"—from students completing their first year of training. I then assessed the relative frequencies in these texts of 986 stylistic features.[6]

They included function words (common words with low semantic content, such as prepositions and forms of modal and auxiliary verbs), parts of speech (such as nouns, verbs, and prepositions), and common part-of-speech bigrams and trigrams. A POS-bigram is a two-token window into the text, showing how frequently ordered pairs of parts of speech appear in the text. (Section 3.3 explains these concepts in greater detail and shows examples.)

I then performed two classes of analysis to parallel those performed in the Argamon/Koppel 02/03 study. Those analyses are described in Section 4.4. Statistical analysis shows whether the features calculated in the previous step were statistically different for men and women and whether men or women were more likely to use the feature. Machine-learning analysis allowed me to consider whether subtle differences that were not prominent enough in this sample to be statistically significant might nevertheless identify patterns of difference prominent enough to permit a machine-learning algorithm to classify the texts.

Chapter 5: Study findings

After carefully carrying out the design described in Chapter 4, I assessed the findings. Neither the statistical analysis nor the machine-learning analysis delivered results that suggested any pattern of difference in the stylistic choices made by the Gender F

[6] As I explain in Chapter 4, I began with 1,074 features that are explained more fully there; I chose to exclude four parts of speech, and 84 of the function words on my list did not appear in my corpus. Thus I actually assessed 986 features.

and Gender M authors. There were some statistically significant differences, and some machine-learning algorithms were able to classify texts by author gender fairly successfully when considered against some previous studies. But the features that differed significantly and those that proved mostly useful in machine-learning classification did not fall into any pattern of Gender F and Gender M preference.

The findings from the statistical analyses appear in Section 5.2. I first compared them particularly to Argamon, Koppel, Fine, and Shimoni (2003), but second I examined them overall. Argamon, Koppel, Fine, and Shimoni (2003) had found statistical differences placing men on the informational end and women on the involved end of the involved/informational dimension first described by Biber (1995). The present study did not show the statistical differences that supported a Gender F/involved and Gender M/informational dimension within this writing sample that Argamon, Koppel, Fine, and Shimoni (2003) had found in theirs. Second, after examining the statistics overall, I also conclude there was no meaningful pattern of stylistic differences between Gender F and Gender M writers.

Section 5.3 presents the findings from the machine-learning algorithm (MLA) trials. Generally, trials of MLAs using all 986 features studied in this project did not yield results that were statistically better than randomly assigning genders to texts. After performing feature selection, a process by which the features that improve the performance of the MLAs are singled out and other features are discarded, I was able to obtain observed agreement between the texts' author-gender and the MLA-assigned gender at levels statistically better than chance. Because published studies generally do not report observed agreement below 66%, I arbitrarily set that is my threshold of practical significance. Three MLAs were able to classify texts accurately enough to be practically significant. The best performance was 73.19%. This is lower than that achieved by Koppel et al. (2002), but it is still as high as many other published studies. However, when examining the features that allowed the MLAs to perform well, it became apparent again that there was no real pattern of variation between texts by the Gender M and Gender F authors: the features that proved useful to one MLA were generally not useful to another MLA.

Chapter 6: Discussion of study findings

Chapter 6 explains the findings in this empirical study using cognitive pragmatic rhetorical (CPR) theory. It first provides a brief overview of CPR theory contextualized into this study. It then expands the discussion of CPR theory to describe a CPR-theoretic production and interpretation process. It then considers how CPR theory accounts for gendered language and students' genre knowledge. Finally, it offers possible explanations of the findings in this study.

Section 6.2 provides an overview or reprise of the concepts of CPR theory previously set out in Chapter 2. This time though, CPR theory is contextualized among the participants in this study, considering their goals and assumptions and their beliefs about their instructors' goals and assumptions. That section extends the earlier description of CPR theory by explaining the CPR-theoretic production and interpretation procedures. Section 6.3 places CPR theory in the context of the present empirical study. Section 6.4 shows that CPR theory accounts for gendered language. And Section 6.5 shows that CPR theory accounts for genre knowledge. Finally, Section 6.6 shows how CPR theory may account for the findings in the present study.

Chapter 7: Conclusion

Chapter 7 brings this dissertation to a conclusion by describing limitations of the empirical study and then discussing implications both of the study in particular and of CPR theory in general. Finally, it considers questions that may motivate future research.

Appendices and the University Digital Conservancy

More than a third of the bulk of this dissertation consists of appendices. These documents describe aspects of the empirical research study I performed and provide detailed reports of all the statistical and machine-learning analyses discussed below. I will mention those appendices in the text as their contents may become useful to the reader.

But a much greater wealth of information and data is available through the University of Minnesota's University Digital Conservancy. I have deposited with the UDC extensive data files, including all the original writing samples (after anonymization) I used in this study and the XML files resulting from manual annotation of them. Other researchers may make use of these materials either to verify the efforts in which I've engaged here or to extend them. All these materials are described in Appendix A. These materials should be discoverable on the Internet with a search engine search including "University of Minnesota," my name, and the title of the dissertation. Note, though, that they will be embargoed until May 2017 so that I can make the effort of developing publications from this dissertation before these materials are widely available.

Chapter 2

Cognitive pragmatic rhetorical theory: A framework

2.1 Introduction

The purpose of this chapter is to propose and justify the use of what I am calling "cognitive pragmatic rhetorical (CPR) theory." CPR theory takes account of a Writer's[1] goals and beliefs about the world and explains the efforts the Writer employs in finding, discovering, or inventing her communicative performances with the principle of relevance, which holds that she will expend effort in her writing choices that is commensurate with the strength of the goals for which she writes and the effect she expects her writing to have on the cognition of the Readers(s). In other words, relevance is a ratio of effect to effort.

This chapter describes my adaptation of the relevance theory of Sperber and Wilson (1995), which I will refer to as SWRT, with important theoretical modifications supplied by Straßheim (2010), grounded in the work of Alfred Schutz (Schutz, 1964, 1966, 1973) (which I will refer to as SSRT); and with modifications that I have provided myself. The final framework is what I refer to as *cognitive pragmatic rhetorical (or CPR) theory.*

[1] I employ the practice of referring to the hypothetical "Speaker" and "Hearer" with an initial capital, using feminine pronouns for the former (Speaker=she) and masculine pronouns for the latter (Hearer=he). I will occasionally refer to the Writer (also with feminine pronouns) and the Reader (with masculine pronouns), but you should regard Speaker and Writer as interchangeable; I will interchange Hearer and Reader as well.

This chapter cannot provide a comprehensive review of any of these theories, but it does describe a theoretical framework useful for understanding the empirical study in this dissertation described in Chapters 4 and 5; I finish supplying the CPR framework and apply it to the findings from that study in Chapter 6.

CPR theory provides a framework for analysis not available in any of these antecedent theories, in that it takes account of the Writer's cognitive environment: her accessible goals and assumptions (propositional beliefs about the world) and the extent to which she is committed to them; her emotional state; the cognitive environment(s) she imputes to her Reader(s), including any expectations the Writer thinks the Reader has about the Writer's performance; and the Writer's express and tacit knowledge about likely cognitive effects of her performance on the Reader(s). The name "cognitive pragmatic rhetorical theory" suggests a grounding in three different disciplines. Each of them—pragmatics, rhetoric, and cognitive science—contributes to understanding communicative performances. But principally, I see CPR theory as a bridge between rhetoric and linguistic pragmatics, particularly the relevance theory of Sperber and Wilson (1995).

The suggestion to build a bridge between pragmatics and rhetoric immediately raises a question: Why build it? In Section 2.2, I'll explain my metatheoretical commitments and motivations. I will then introduce what I call *classical pragmatics* in Section 2.3 and identify its shortcomings for my project, particularly insights it fails to incorporate from rhetoric and cognitive science.

In Section 2.4, I introduce the relevance theory of Sperber and Wilson ("SWRT") which I see as the foundations and main cables of a bridge between rhetoric and pragmatics. I briefly describe SWRT's central principles and mention how they overcome some of the shortcomings of classical pragmatics. Of particular interest is the concept of *relevance* as a ratio of positive cognitive effects to cognitive effort that undergirds human cognition. SWRT nevertheless leaves some problems of classical pragmatics unresolved.

I then describe the extension of Sperber and Wilson by Straßheim, founded in the concepts of "relevance" and "typification" from Alfred Schutz ("SSRT"). I see Straßheim's work as providing the suspension cables of the bridge. I present its valuable additions to SWRT, particularly the notion that an agent's goals figure into her assessments of relevance. But note that SSRT, too, leaves some shortcomings of antecedent theories unremedied.

Finally, what remains is to deck the bridge and open it for traffic, which I do with CPR theory in Section 2.5. There, I explain the components of the cognitive environment and the concept of imputed cognitive environment. I'll allude there to the CPR-theoretic production and comprehension procedures, which I'll describe in more detail Chapter 6.

Like SWRT, CPR theory rejects Grice's communicative principle as neither necessary nor sufficient to explain human communication. Like SSRT, CPR theory takes into account that communicative agents may assess relevance by reference to more than just the propositional content of communications; emotional states and goals of communicative agents play a critical role in their production and interpretation of utterances. Speakers' goals may prove particularly critical for explaining the findings in the study in this dissertation. CPR theory extends SSRT by spelling out an utterance-production procedure and taking account of the unconscious operations, like habitual performances, that influence utterance production and interpretation.

In Chapter 3, I suggest how CPR theory accounts for the research construct of genre, and in Chapter 6, I show how CPR theory may account for the results in the empirical study described in Chapters 4 and 5. Finally, in Chapter 7, I consider the limitations and implications of this study and of CPR theory, and I suggest some directions for future study.

2.2 Metatheoretical concerns

I call what I am proposing in this chapter "cognitive pragmatic rhetorical theory" or "CPR theory." Before talking about CPR *theory*, I should take a moment to describe my own *meta*theoretical commitments, which is to say my "underlying beliefs which generate a particular approach" or "ideology or theoretical presupposition" (Figueroa, 1994, p. 4). I will then explain why the theory meeting these metatheoretical commitments should bring together rhetoric, pragmatics, and cognitive science.

2.2.1 Epistemic commitments

For me, CPR theory fits into a paradigm that Lazaraton (2003) described as "quasi-foundational." Though Lazaraton was exploring evaluative criteria for qualitative research, and she did not use the term "metatheory" for the orientations she described, they are nevertheless metatheoretical on Figueroa's definition. Lazaraton (p. 7) considered three possible orientations: In *foundationalism*, there "should be one set of criteria for any kind of scientific research (i.e., positivism/rationalism: the criteria of reliability, internal and external validity, and objectivity)." In *quasi-foundationalism*, there "should be criteria unique to qualitative research (i.e., postpositivism, constructivism), which value (a) theory generation, (b) empirical grounding and scientific credibility, (c) generalizable/transferable findings, and (d) internal reflexivity, in that researcher, context, and effects are taken into account." And finally, in *nonfoundationalism*, "[a]ll such criteria should be doubted and none should be privileged (i.e., postmodernism); altogether new criteria are needed, in as much as judging inquiry is a practical, political, moral affair, not an epistemological one (i.e., poststructuralism)."

I propose CPR theory then as quasi-foundational. Its claims should be grounded in empirical study, studies should point to claims that obtain or can obtain scientific credibility and cautious generalizability (at least across some scale), and researchers in the space should explore their own positions as well as other contextual factors in the research. The claims of CPR theory would thus be subject to argumentation that exhibits "critical reasonableness" (van Eemeren & Grootendorst, 2004, p. 16), according to conventions developed within the field of CPR theoreticians.[2]

In the CPR theoretical model, it should be possible to test for the existence of a phenomenon predicted by theory or to explain a phenomenon shown to exist. For example, in rhetorical theory since Aristotle, polysyndeton and asyndeton have been claimed to have a variety of effects. Oates and Enquist (2006) used examples to argue that asyndeton minimizes the elements in a series while polysyndeton emphasizes them (pp. 698-99). Generally, rhetorical theorists do not feel compelled to answer the empirical questions, for example, whether polysyndeton and asyndeton actually have the effects their theories say they do. Further, they offer no explanations for how it is that

[2] If there ever is more than one.

these two figures function to have the effects claimed for them. I would seek to fill these gaps by empirical means, including observational and (quasi)experimental studies.

In order for CPR theory to be successful, it should function as a framework that can be used to identify variables for characterizing human performances across the communicative landscape; variables that empirical researchers can operationalize for observational, quasi-experimental, or experimental studies of human communication. It should explain what communicative agents are doing and make predictions about what they will do in certain environments.

The name I have chosen for this framework—cognitive pragmatic rhetorical theory—suggests a question I will take up in the next section.

2.2.2 Why "rhetorical," "pragmatic," and "cognitive"?

In my view, rhetoric is a process of changing minds using language, paralanguage, and text—broadly defined. The Speaker produces an utterance or text directed to the Hearer in order to change the Hearer's state of mind, changing what he believes about the world, altering his emotional state (including affecting the way he feels toward the Speaker), redirecting his goals, or most likely some combination of these.

Pragmatics and rhetoric have much in common. Pragmatics has been described as "the study of language from the point of view of the user, especially the choices [s]he makes, the constraints [s]he encounters in using language in social interaction, and the effects [her] use of language has on the other participants in an act of communication" (Crystal, 1985, pp. 278-79). This connects naturally with the definition of rhetoric I provided in the previous paragraph. What's more, there is a model within the field of experimental pragmatics for understanding language use that seems particularly compatible with rhetoric: the relevance theory of Sperber and Wilson (1995). Relevance theory serves as the foundation and provides the main cables for a bridge between rhetoric and pragmatics. As the theory I propose draws heavily on relevance theory and pragmatics, "pragmatics" belongs in its name.

My own interests and commitments derive from cognitive science in a variety of ways. But by "cognitive science," I do not mean the first-wave artificial-intelligence models of cognitive science, which viewed the brain as a sort of computer, operating via rational inference on symbols (Newell & H. Simon, 1976). Nor do I see myself as firmly entrenched in the camp of any of the subsequent models of cognition, including connectionism (Churchland, 1989), embedded/embodied cognition (Clark, 1997), or dynamical systems theory (Juarrero, 1999). Rather, I'm committed generally to the principle embraced by all these models that there is a sytematicity to human cognition that is worth exploring and understanding.

My metatheoretical, epistemic, and disciplinary commitments thus describe the inputs for CPR theory and suggest means for assessing its success. Before laying out its principles, though, we need to consider the antecedent contributions of classical pragmatics and relevance-theoretic models of human communication.

2.3 "Classical" pragmatics, rhetoric, and cognition

"Everyone agrees that pragmatics is concerned with the study of meaning" (Chapman, 2011, p. 1). It is "the study of language from the point of view of the user, especially the choices [s]he makes, the constraints [s]he encounters in using language in social interaction, and the effects [her] use of language has on the other participants in an act of communication" (Crystal, 1985, pp. 278-79); and it "has to do with all context-dependent aspects of meaning 'systematically abstracted away from the pure semantics of logical form'" (Nerlich and Clarke (1996, p. 4), quoting Horn (1992)).

This section describes classical pragmatics and its shortcomings that CPR theory addresses. Section 2.3.1 provides the overview of classical pragmatics. As pragmatics is associated with the choices that Writers make, it seems a natural place to start when seeking a theory to account for Writers' choices in the study in this dissertation. Section 2.3.2 explains the ways in which pragmatics and rhetoric interact (or fail to do so), noting that classical pragmatics does not account for emotions. More importantly for this study, pragmatics does not account for the fact that Speaker and Hearer may not have the same goals for the communicative interaction or for the fact that the goals of the Speaker may play a significant role in her selection of communicative performances. Pragmatics also lacks the insight of rhetoric that communicative interaction is not only (and perhaps not principally) about sharing information, but is directed at persuasion. Classical pragmatics fails to account for important discoveries from cognitive science and cognitive psychology, particularly insights that rational inference, which is at the heart of classical pragmatics, must share the stage in communication with cognitive biases; Section 2.3.3 explores those issues. In Section 2.4, we move on to consider whether relevance-theoretic models of pragmatics resolve these problems.

2.3.1 Overview of classical pragmatics

The prevailing theories of language-in-use owe much to work in the philosophy of language by Austin (1975), Searle (1970, 1979), and Grice (1989). Taken together, these scholars can be viewed as the founders of linguistic pragmatics; much of the research in that field is grounded in these works or at least sets itself up in response to them. I will refer to their program as *classical pragmatics*. Classical pragmatics introduced into philosophical and linguistic discourse the notion that meaning cannot be found merely by decoding sentences based on the semantics of words and the manner of their combination (syntax). Rather, scholars of pragmatics asserted that meaning must be derived from the application of inference to the products of decoding.

So, pragmatics generally contrasts the *content of a sentence* (the domains of syntax and semantics) with the *context of an utterance* (the domain of pragmatics). An example may be helpful. Assuming we have two speakers of French, sentence (1) appears to be a literal response to (2).

(1)	La	plume	de	ma	tante	est	sur	le	bureau	de	mon	oncle.
	The	pen	of	my	aunt	is	on	the	desk	of	my	uncle.
	'My aunt's pen is on my uncle's desk.'											

(2)	Où	est	la	plume	de	ma	tante?
	Where	is	the	pen	of	my	aunt?
	'Where is my aunt's pen?'						

Assuming the utterers of these two sentences are siblings who speak French, that the utterer of (2) is seeking the pen of their common aunt, and the utterer of (1) intends to inform him that the pen in question is on their common uncle's desk, all that is required for comprehension of (1) by the utterer of (2) is to *decode* it.[3] By complicating the context, however, we can complicate the interpretation of (1).

[3] Let's ignore for the moment that he must resolve the reference *my* as some person and *my aunt's pen* as a particular possession of that person.

Sentence (1) has reputedly appeared in low-quality commercial language phrasebooks and has been described as practically useless ("La plume de ma tante (linguistics)," n.d.). Imagine that Bill, a linguist, and Anne, a teacher of German, are professionally familiar with these facts, and that each expects the other is as well. They have the following exchange:

(3) *Bill:* What do you think of Smith's new German textbook?

(4) *Anne:* La plume de ma tante est sur le bureau de mon oncle.

Here, here Bill cannot hope to interpret Anne's utterance by *decoding* (4). Even if Bill speaks French, the propositional content of (4) does nothing to answer his question. Nevertheless, the context and Bill and Anne's knowledge of each other and their fields is likely to guide him to a correct interpretation—perhaps as soon as he recognizes the phrase after the first few words—or to ask her for clarification perhaps in the form of this further exchange:

(5) *Bill:* Superficial and useless, I take it?

(6) *Anne:* Yup.

Here, decoding provides little if any interpretive purchase. Pragmatics, then, busies itself with going beyond code models of language, usually by application of *inference*. According to Grice (1989), an utterance may encode a speaker's meaning, or it may merely be evidence of her meaning. For Grice (1989, p. 26), "talk exchanges are characteristically, to some degree at least, cooperative efforts—[with] a common purpose or set of purposes, or at least a mutually accepted direction." Grice famously offered his cooperative principle and maxims deriving from it:

(7) **Grice's Cooperative Principle (CP):** Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged (Grice, 1989, p. 26).

The conclusions that the Hearer draws from the inferential process are what Grice calls "implicatures."

Classical pragmatics is nevertheless still focused on decoding and rational inferences about meaning implicatures. Common to the theories of Searle (1970, 1979) and Grice (1989) is the principle that the Hearer should first decode the literal meaning, or "what is said," and then use inference (relying on the CP and the maxims) to derive the implicatures only where the decoded meaning would fail to satisfy the CP. That is, they assume that sentence meaning is close to Speaker meaning and that inference bridges the gap. Concluding that inferences to derive Speaker meaning may be conventional and formalizable, they seek rules for transformations (similar to those in generative syntax) to support the inferences.

Two things are important to note for later discussion: First, for Grice and many theorists of pragmatics in the classical mode, implicatures are propositions about the world. That is, the goal of inferential comprehension by the Hearer is that he will believe a larger set of (hopefully true) propositions about the world. Thus, after the exchange between Bill and Anne in (3) and (4), Bill could entertain a thought or *representation* with the content in (8), or at least the content in (9).[4]

(8) 'Smith's new German textbook is superficial and useless.'

(9) 'Anne thinks Smith's new German textbook is superficial and useless.'

As Dascal and Gross (1999) and Liu and Zhu (2011) (discussed in Section 2.3.2) explain, rhetoric is concerned with the *effects* (at least persuasive, but perhaps others) of communication, not merely *propositional content*. I shall address this concern in Section 2.5.

[4] The concept of *representation* as a container that holds propositional content is complicated in cognitive science. Grush (2001) described the contest of the original "cognitive revolution"—which rejected behavioralism—and the "cognitive counter-revolution"—which rejected the computational-symbolic views of the cognitive revolution:

> [T]he counter-revolution is right that representations understood as symbols structured along something like first-order predicate logic and manipulated via something like inference rules probably have very limited application in understanding the various aspects of cognition; but the counter-revolution is quite wrong to try to exorcise the notion of representation altogether. Representation is here to stay. How to correctly understand its various manifestations is what is up for grabs.

For the moment, I will accept the notion of representation as a container of propositional content without further comment, but I will discuss below that I believe emotions and goals or desires fall outside that conceptual boundary.

Second, these theorists consider figurative language (including metaphor and loose uses) exceptional, often finding that it "flouts" the maxims and requires a special comprehension procedure. The relevance theory of Sperber and Wilson (1995) partially addresses this concern, as we shall see in Section 2.4.1.

2.3.2 Classical pragmatics and rhetoric

To the extent that pragmatic theories extend language study away from code models (Bazerman, 1988, pp. 299-300), they are likely compatible with rhetoric. Indeed, rhetoric and pragmatics have ancient history in common: In their history of early pragmatics, Nerlich and Clarke (1996, pp. 9-10) found its roots in Protagoras' identification of the moods or modes of speech ("statement, question and imperative") and in Aristotle's *Rhetoric* (Aristotle, 2007). Rhetoricians and some critical theorists of interest to them are also deeply interested in the work that words *do*, regardless of what their *literal meaning* (if there is such a thing) might be. Indeed, many theorists in rhetoric, argumentation theory, and critical theory have employed concepts from classical pragmatics. Here is just a small sample: Berkenkotter and Huckin (1994) used Searle's speech act taxonomy in their analysis of text genre. Bhatia (1993, pp. 45-75) used Grice to aid in interpreting business communication. Argumentation theorists van Eemeren and Grootendorst (2004) adopted a modified Gricean framework and Searlian taxonomy for analyzing argument, and Jacobs (2000) emphasized the "pragmatic force" of argumentation. Finally, Butler (1997) critiqued and adapted Austin's work to develop the concept of "performativity."

But classical pragmatics has proven somewhat difficult to reconcile with rhetoric. Dascal and Gross (1999) proposed a "marriage" of pragmatics and rhetoric. They advocated for what they described as a "cognitive theory of rhetoric" (p. 108), but what I would refer to as a *rationalist* theory of rhetoric. They identified at least three reasons why other scholars might not hold their peace at the nuptials of these two fields: First, Gricean pragmatics is dialogic; Grice discusses two interlocutors, the Speaker and the Hearer. Rhetoric must account for a Speaker with an audience or a Writer composing a text perhaps without a clear awareness who her audience is. Second, rhetoric is generally not seen as having rational inference at its center. Though one of Aristotle's artistic proofs, *logos*, is clearly consistent with the concept of rational inference, Dascal and Gross acknowledged that most readers would not see *pathos* and *êthos* in the same light. Third, Grice's Cooperative Principle (CP) does not seem to permit misdirection, a common tool in communication and one commonly analyzed by rhetoricians.

Dascal and Gross (1999) addressed these concerns: First, they argued that the dialogic nature of Gricean pragmatics is not incompatible with the "continuous interchange" of oratory, where the "audience will not be slow to indicate to [the Speaker] that the persuasive situation no longer pertains" (p. 109). Second, they argued that rational inference could be applied to *pathos* and *êthos* in determining their evocative effects rather than propositional content. And third, they proposed modifying the CP so that it states that only the appearance of cooperation must be maintained.

Liu and Zhu (2011) asserted that Dascal and Gross misdiagnosed a mismatch between pragmatics and rhetoric. They argued that pragmatics has "foundational principles," such as the CP (7), but that rhetoric is identified by "modus operandi" such as *pathos*, *êthos*, and style (p. 3408). Liu and Zhu (2011) challenged the very rationale for formalizing relations between rhetoric and pragmatics and asked "Can we lift a key term from one discourse and transplant it to another without taking into account the overall relatedness or affinity between them?" (p. 3406). They proposed rhetoric and pragmatics as *antistrophoi*, echoing Aristotle's juxtaposition of rhetoric and dialectics under the same moniker. Consequently, they saw the CP and Grice's maxims as requiring *antistropoi* within rhetoric. They proposed the first of these—what they called the "foundational principle in rhetoric"—the Non-Cooperative Principle or NCP:

(10) **Non-Cooperative Principle (Liu & Zhu, 2011):** When engaging and addressing another party, always proceed without assuming that the addressee would voluntarily cooperate with you in producing the effect or result you desire. When being engaged and addressed, always respond without assuming that the addressor would voluntarily cooperate with you in your effort to see the matter concerned from your own perspective or to reach a decision about it on your own terms.

It is difficult, however, to see the NCP in (10) as *antistrophos* to the CP in (7); instead, they appear to be *antitheses*. Grice's entire system relies upon the CP for any coherence that it achieves; and yet the NCP entirely undermines it. And what kind of foundational principle is the NCP? It appears in the form: Always proceed or respond without assuming that the other party is being cooperative. Resisting an assumption about whether a particular fact about a talk interaction is true is hardly a foundational principle.

Liu and Zhu (2011) and Dascal and Gross (1999) did contribute at least three important insights into the relation between rhetoric and pragmatics. First, rhetoric overtly accounts for emotion in understanding human communication (or at least persuasion). At least since Aristotle, rhetoric has recognized emotion as a source of persuasion. Dascal and Gross (1999) proposed that pragmatics and rhetoric could be joined by applying rational inference (what they called "rhetoric as a cognitive theory" (p. 108)) to *pathos* and *êthos*. I will address that with my own suggestions in Section 2.5. Second, the Speaker and Hearer may have different goals, and Liu and Zhu (2011) claimed that rhetoric has long taken account of that fact. Straßheim (2010) adapts the conception of *relevance* from Schutz (1973, 1964, 1966) to address this concern, as we shall see in Section 2.4.2. Third, classical pragmatics is grounded in a presumption that the purpose of communication is the sharing of information; rhetoric explores persuasion. My adaptation of the relevance theories of Sperber and Wilson (1995) and Straßheim (2010), described in Section 2.5, addresses these concerns.

2.3.3 Classical pragmatics and cognitive science

But before moving to a consideration of various relevance theories, it is worth noting that classical pragmatics is at odds in certain respects with contemporary cognitive science, which suggests that humans (like many other animals) do not routinely engage in the kind of rational inference posited by the founders of classical pragmatics and that their inferences are governed by emotional forces. According to Gigerenzer and Brighton (2009), the human mind evolved as a toolbox, a set of capacities, to survive in certain environments. Among the important implements in our cognitive toolboxes are *heuristics*, "fast and frugal" (Gigerenzer & Brighton, 2009, p. 109) techniques for drawing conclusions or inferences about our environments that are not necessarily rational in any substantive sense. For example, Gigerenzer and Brighton (2009) identified several key heuristics, including the recognition heuristic (if one of two alternatives is recognized, it is preferred); and the fluency heuristic (if two alternatives are recognized, but one is recognized more quickly, it is preferred).[5] Such efforts to minimize cognitive effort are outgrowths of what H. A. Simon (1990) called "bounded rationality"—there is only so much cognitive effort that a brain can bring to bear on survival problems. Heuristics function as a kind of "procedural rationality" or "ecological rationality," provided they are used in the environments in which they evolved. Heuristics do not always work well, of course, and Kahneman (2003) has pointed to some of the poor results that can arise from them.

Heuristic, ecological rationality stands in contrast to "the canonical definition of rational inference as weighting and adding of all information (as long as it is free)" (Gigerenzer & Brighton, 2009, p. 113). Clark (1997) referred to this as "compositional analysis"—conscious problem-solving that applies computation to the complete set of available evidence. Of course, obtaining "all information" is never free, and neither is the effort required to process it (Gigerenzer & Brighton, 2009). In fact, rational inference may not be the most effective approach for making the right decision, especially where predictions about future events are concerned; from a survival standpoint, most organisms have a greater need for predicting future events than for explaining past events.[6] Gigerenzer and Brighton (2009) described several studies where they compared the success in making predictions of techniques that weigh all available data (like linear and multivariate regression) against heuristic decision-making methods (some performed using machine-learning algorithms). They found that heuristic predictions were almost always as accurate or more accurate than those made using rational inference.

Given the important role of language in human history (and probably pre-history) and the rich social practices of other primates, humans likely evolved with rich heuristics

[5] Gigerenzer and Brighton's recognition heuristic is akin to the concept of the affordance proposed by Gibson (1979). Affordances make good evolutionary/cognitive sense: It's much more efficient to develop a heuristic for recognizing certain patterns of characteristics according to their utility than to have to perform a compositional analysis of separate characteristics to come to the same conclusion each time a new environment is acquired. For example, a bird might recognize an opening of a certain size on a certain type of surface as a nesting spot without calculating the size of the opening or assessing the characteristics of the surface more than a few feet beyond the opening; this is why a bluebird might recognize a nesting box placed by a human as a nesting place, despite the fact that the structure overall looks nothing like the bird's natural nesting places. I'm grateful to Dr. Carol Berkenkotter for this example.

[6] Humans may be an important exception here, and I am certainly not suggesting that important social decisions about the future be made without reference to rational inferences from data we have about topics like climate crisis, for example.

specifically for social interaction. These heuristics predate and frequently supervene decision-making in the rational inferential mold (Haidt, 2001). Mercier and Sperber (2011) went a step further, and argued that *human reasoning evolved in response to the need to persuade other humans*, that reasoning evolved for the "production and evaluation of arguments in communication" (Mercier & Sperber, 2011, p. 58). They claimed that reasoning evolved in service and support of decisions already made, perhaps largely by non-rational processes.

Two other issues suggest that classical pragmatics can provide at most an incomplete picture of human communication and the meanings that humans interpret in their communication: classical pragmatics' view of metaphor and its failure to account for unconscious "calculations" in language use. As an example of the former, consider the well known study of metaphor by G. Lakoff and Johnson (2003), according to which metaphors play a central role in structuring our understanding of language. For G. Lakoff and Johnson (2003), the figurative meaning of metaphors cannot be separated from the "literal" meanings. By contrast, classical pragmatics proposes that the Hearer calculate literal meaning first and consider the possibility of metaphorical meaning only where literal meaning would result in a breach of the CP or a maxim. As an example of the latter issue, consider Fausey and Boroditsky (2010), who showed that non-agentive descriptions of events caused listeners to conclude that involved actors had less culpability. They showed, for example, that listeners who heard (11) were more likely to view Timberlake as culpable than listeners who heard (12), even when the latter were aware from contextual information that Timberlake did the ripping.

(11) Timberlake ripped the costume.

(12) The costume ripped.

Classical pragmatics thus does not account for the impact of metaphor and unconscious processing or heuristics in determining utterance meaning.

In sum, classical pragmatics relies upon a sort of rationalist assumption about human conduct, that humans use rational inference to interpret communication and that the object of their interpretations is propositional content. The effort of rational inference or compositional analysis requires much cognitive energy, and the human mind has evolved to minimize cognitive effort, just as the human body evolved to minimize physical effort. Human communication should be no different in this regard than any other human activity. And human language and cognition are subject to a wide variety of heuristics and cognitive biases that suggest substantive rationality does not govern utterance interpretation in any event.

Among the participants in the study in this dissertation, we must account for the fact that they probably make gendered habitual stylistic choices when producing informal utterances, but they managed their stylistic choices in their law-school writing in such a way as to avoid a gendered style. My intuition was that unconscious habits of production and the balance of cognitive effort and effect played a role in this outcome. These are among the concerns that Sperber and Wilson (1995) attempted to address with "relevance theory."

2.4 Relevance-theoretic pragmatics

Beginning in the 1970s, linguist Deirdre Wilson and anthropologist and cognitive scientist Dan Sperber began exploring pragmatics in their work. By the early 1980s, they were working together to address concerns they had with the classical pragmatic model. In 1986, they published the first edition of *Relevance: Communication & Cognition*; the second edition followed in 1995 (Sperber & Wilson, 1995). Since its introduction, Sperber and Wilson's Relevance Theory (SWRT) has become a mainstream theory in linguistic pragmatics, particularly in experimental pragmatics, which employs empirical methods rather than philosophical speculation for pragmatic theory building. As of January 2015, Google Scholar reported more than 14,000 citations to Sperber and Wilson (1995). Yus (2015) provided a massive, thematically categorized bibliography of citing works.

I contend that SWRT provides a considerably more *rhetorical* model of pragmatics than classical pragmatics, and it addresses some of the concerns with classical pragmatics described in Section 2.3. SWRT has nevertheless found little welcome in the field of rhetoric and technical communication. This may be due in part to antipathy expressed by Dan Sperber toward the rhetorical tradition. For example, in Sperber and Cummins (2007) he wrote in a footnote crediting influences: "I will name no rhetoriticians [sic] here, for it was precisely to free myself from their influence that I undertook this work." Harris (2007, p. 359) quoted Sperber and Wilson as saying "If relevance theory is right, then. . . rhetoric has no subject matter to study, or to teach."

Nevertheless, I claim in Section 2.4.1 that SWRT goes far toward bridging the gap between rhetoric and classical pragmatics; SWRT might be seen as providing the foundations and main cables of the bridge. Section 2.4.2 describes modifications to SWRT proposed by Straßheim (2010) and which I call Schutz-Straßheim Relevance Theory or SSRT; these additions, together with the cognitive principles described in Section 2.3.3 might be seen as the bridge's suspension cables. I hope in Section 2.5 to deck the bridge and open it to traffic.

By crossing this bridge, we can take account of a Writer's goals and beliefs about the world and explain the efforts the Writer employs in finding, discovering, or inventing her communicative performances. Central to this effort is the principle of relevance, which holds that she will expend effort in her writing choices that is commensurate with the accessibility and strength of the goals for which she writes and the effect she expects her writing to have on the cognition of the Readers(s). In other words, relevance is a ratio of effect to effort. This ratio will permit us to understand how and why the participants in this study invested the effort to abandon gendered communicative habits, if indeed they had them to begin with.

2.4.1 SWRT: The relevance theory of Sperber and Wilson

The relevance theory of Sperber and Wilson (SWRT) is proposed as a cognitive theory of language-in-use. Sperber and Wilson (1995) claimed that the effort to comprehend an utterance is balanced against its positive cognitive effects. The more positive cognitive effects a stimulus has, the more relevant it is; the more difficult to process a stimulus is, the less relevant. It is this ratio that Sperber and Wilson (1995) refer to as "relevance" in the context of human communication: The greater the utility and the lesser the effort to comprehend, the greater the relevance (see also Wilson & Sperber, 2006, 2012). They argued that we maximize relevance because of the way our brains are evolved (Wilson & Sperber, 2006, p. 610), and that "the search for relevance is a basic feature of human cognition, which communicators may exploit" (Wilson & Sperber, 2006, p. 608).

According to Wilson and Sperber (2006), "[t]he central claim of relevance theory is that the expectations of relevance raised by an utterance are precise and predictable enough to guide the Hearer toward the speaker's meaning" (Wilson & Sperber, 2006, p. 607). "The universal cognitive tendency to maximize relevance makes it possible (to some extent) to predict and manipulate the mental states of others" (Wilson & Sperber, 2006, p. 611). In other words, it is not necessary to assume or accept Grice's Cooperative Principle. In fact, Sperber and Wilson (1995) expended considerable effort to show that Grice's CP is neither a necessary nor a sufficient condition for communication and comprehension.

Wilson and Sperber's theory is more fully developed in extended works on the subject (Sperber & Wilson, 1995; Wilson & Sperber, 2012). An accessible and useful summary of SWRT appears in Wilson and Sperber (2006). But of central importance for the empirical study in this dissertation are several concepts that I will discuss briefly here: the cognitive and communicative principles of relevance; the relevance-theoretic comprehension procedure; "explicatures"; and SWRT's stance toward metaphor and irony.

Underlying all of SWRT, then, are the two principles of relevance:

(13) Cognitive (First) Principle of Relevance: Human cognition tends to be geared to the maximization of relevance. (Wilson & Sperber, 2006, p. 610)

(14) Communicative (Second) Principle of Relevance: Every ostensive stimulus conveys a presumption of its own optimal relevance. (Wilson & Sperber, 2006, p. 612)

The cognitive principle in (13) governs all human cognition. The communicative principle in (14) comes into play in human communication.[7]

Wilson and Sperber (2006, p. 612) proposed that a Hearer confronted with an ostensive communicative stimulus employs the relevance-theoretic comprehension procedure:

(15) Follow a path of least effort in computing cognitive effects: Test interpretive

[7] By *ostensive* communicative stimulus, Wilson and Sperber refer to one "designed to attract an audience's attention and focus it on the communicator's meaning" (Wilson & Sperber, 2006, p. 611). I won't take up this issue here for the moment, but it has interesting implications for any philosophy of communication.

hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.

(16) Stop when your expectations of relevance are satisfied (or abandoned).

Note that for Wilson and Sperber, the effect side of the relevance calculation, represented by “cognitive effects” in the comprehension procedure, is measured by an increase in the Hearer’s store of assumptions—that is, by the true representations about the world that can be stated in propositional form and that the Hearer can infer from the Speaker’s utterance. This constraint ignores the fact that some assumptions will be more important to the Hearer than others, depending on what his goals are for the communicative interaction. We will take this up in Section 2.4.2. What’s more, this constraint ignores important aspects of everyday communications, like phatic communications intended not to communicate information but to enhance the affiliation between Speaker and Hearer or to create an emotional response in the Hearer that cannot easily be summarized in propositional form. We will take this up in Section 2.5.

In any event, under the relevance-theoretic comprehension procedure, the easy interpretation should be the first; and by definition, it will be the best (though it might not be correct). This approach is consistent with the cognitive heuristics called “take-the-best” and “satisficing” by Gigerenzer and Brighton (2009, p. 130).

Grice (1989) contrasted that portion of interpretation consisting of decoding “what is said” with the recovery of “implicatures” by means of inference. Imagine that Laura and Marco are sitting in a room that is cold because a window is open; both of them know the room is cold. Laura utters (17).

(17) *Laura:* It’s cold in here.

If Marco decodes (17) successfully, it tells him nothing new. Under Gricean theory (17) flouts the maxim of quantity (be informative). It therefore licenses Marco to derive implicatures, one of which might be expressed by the command in (18) or the proposition in (19).

(18) ‘Close the window.’

(19) ‘Laura wants Marco to close the window.’

Importantly, Grice (1989) calls for the utterance to be decoded and then the inferential process to apply only where necessary.

Sperber and Wilson, in contrast, claimed that even the decoding process requires inferences about disambiguation of word senses and reference resolution, what they call "explicatures" or "identification of explicit content" (Wilson & Sperber, 2006, p. 615). Importantly, they imagined a series of subtasks of the first step of the comprehension procedure in (15) to happen simultaneously and for the results of the this "online" comprehension procedure to influence each other, listed in (20) through (22).

(20) Identify explicatures through "decoding, disambiguation, reference resolution, and other pragmatic enrichment processes" (Wilson & Sperber, 2006, p. 615).

(21) Construct "an appropriate hypothesis about the intended contextual assumptions (implicated premises)" (Wilson & Sperber, 2006, p. 615).

(22) Construct "an appropriate hypothesis about the intended contextual implications (implicated conclusions)" (Wilson & Sperber, 2006, p. 615).

SWRT takes all these subtasks as requiring inference and as taking place simultaneously. The Hearer does not decode "literal meaning" or "what is said" first and only then apply inference.

One advantage of this approach is that SWRT does not have to regard metaphor and irony as exceptional or as arising only from a clash between or flouting of Gricean maxims. Consider this example from Wilson and Sperber (2012):

(23) Žižek is another Derrida. (p. 109)

Wilson and Sperber claimed that in (23), "'Derrida' is used as a common noun to denote a category of flamboyant and obscure philosophers à la Derrida" (p.109). At no time, they argued, does the Hearer in this case entertain a literal interpretation of (23). Rather, they contend that this is a "category extension," that 'Derrida' here is used to indicate an ad hoc concept, which they call DERRIDA*, that represents the category of which both Derrida and Žižek are members. Wilson and Sperber maintain that the Hearer reaches this interpretation by mutual adjustment of the explicit and implied content in the utterance. This is exactly the same method that the Hearer uses to interpret any other utterance, including "strictly literal interpretations" (p. 108).

SWRT, the relevance theory of Wilson and Sperber, addresses some of the concerns raised in Section 2.3. Notably, it dispenses with the Cooperative Principle of Grice as unnecessary, removing an impediment to the bridge between pragmatics and rhetoric identified in Section 2.3.2. SWRT shifts away from a preference for decoding, still evident in classical pragmatics, and offers decoding and inference as parallel process operating online rather than one before the other. And it accounts for figurative language as a natural part of language production and comprehension, as suggested by cognitive linguistics.

SWRT does not address all the shortcomings of classical pragmatic theory, however. Important for the study in this dissertation is a Writer's production of utterances; SWRT does not account for the Speaker's or Writer's production of utterances at all, and instead focuses only on the Hearer's/Reader's comprehension. It therefore makes little or no effort to explain how it is that the Writer constructs her utterance so as to be optimally relevant to the Reader. SWRT does not address concerns that implicatures in classical pragmatics are propositional in form; rhetoric requires processes that account for persuasive and emotional effects. Though SWRT adopts some of the language of cognitive science, it relies heavily on rational inference, not accounting for all the "fast and frugal" heuristics discussed in Section 2.3.3. For example, though SWRT makes figurative language less exceptional, it does not account for the ways that figurative language and stylistic choices can be used to manipulate the Hearer's unconscious cognition.

This is especially true with regard to habitual performances; a cognitive bias or heuristic called the "fluency heuristic" calls on the agent to choose an option that is recognized faster over one recognized less quickly (Gigerenzer & Brighton, 2009, p. 130). In the empirical study in this dissertation, I potentially need to account for the abandonment of habitual gendered performances participants may have brought to the professional training context. Also important for the study in this dissertation is the impact that the goals of the Writer may have on the process by which she produces her utterance. I must explain how and why is it that the Writer who would probably have produced a stylistically gendered utterance in an informal context instead produces an *ungendered* utterance in response to a high-stakes writing assignment. The work of Straßheim (2010), grounded in theories from Schutz and discussed in the next section, will begin to resolve these problems.

2.4.2 SSRT: Straßheim's extension of relevance theory

The relevance theory of Sperber and Wilson (1995) (SWRT), offers a contemporary reformulation of pragmatics that aligns that field with rhetoric to some extent. But it leaves some matters not satisfactorily resolved. This section takes up revisions to relevance theory proposed by Straßheim (2010) and based on the work of Alfred Schutz, which I refer to as Schutz-Straßheim Relevance Theory or SSRT to contrast it with SWRT. It is especially important for the empirical study in this dissertation that SSRT focuses attention on the Speaker/Writer's production of utterances; on the concept of "typification," which can be seen as a habitual cognitive response to a repeated context; and the consideration of things other than assumptions as constituting cognitive effects, including especially the Writer's goals. SSRT can thus make an account of the writing production of the participants in my study, explaining their willingness to abandon habitual or typified, gendered, communicative performances in favor of the genre or typified form of the court memorandum, all in response to the goal of performing well in a high-stakes environment.

Schutz was a philosopher of sociology, banker, and economist, writing in the mid-20th century (Straßheim, 2010; Bazerman, 2013). Among the objects of his study was human communication and language (Schutz, 1973). Schutz's theory of *typification* has played an important role in contemporary rhetorical and genre theory. For example, Miller (1984) adopted Schutz's typification as a basis for her important contributions to genre theory, writing it "is through the process of typification that we create recurrence, analogies, similarities. What recurs is not a material situation (a real, objective, factual event) but our construal of type" (Miller, 1984, p. 157). Bazerman (2013) devoted considerable attention to Schutz's ideas, especially typification, in Bazerman's comprehensive discussion of a theory of written communication. He described Schutz's conception of typification this way:

> [I]ndividuals, in order to participate in what they see as a meaningful or

> useful social arena, take on what they believe to be principles of that arena. They then use those principles to guide their own behavior and to make their behavior meaningful and intelligible to other participants. Whatever their underlying motives and thoughts are, as the impulses become realized within social action, impulses take on the forms of social types. Those types in turn provide a cognitive orientation for the individual, establishing patterns and principles of thought and identifying relevant knowledge that the individual brings to bear on the circumstances (Bazerman, 2013, p. 68).

Straßheim (2010) argued that Schutz's typification is grounded in the concept of relevance, and that Schutz's relevance-theoretic perspectives on communication complement those of SWRT. Straßheim quoted Schutz: "The concept of relevance is the central concept of sociology and of the cultural sciences" (p. 1413). Elsewhere, Schutz described the relation of relevance to typification:

> [I]t is our passive interest that makes [us] turn toward the object, the interesting object wakening expectations of a particular kind. This is certainly correct, but the term "interest" is simply the heading for a series of complicated problems, which for the sake of convenience shall be called the problem of *relevance*. We turn our interest to those experiences which for one reason or another seem to us to be relevant to the sum total of our situation as experienced by us in any given present (Schutz, 1973, p. 283; emphasis in original).

Straßheim (2010) compared and contrasted the relevance theories of Sperber and Wilson and of Schutz to produce a hybrid, what I call Schutz-Straßheim Relevance Theory or SSRT to contrast it with the SWRT of Wilson and Sperber. In my view, Straßheim made three important contributions: The first was to recognize the importance of explaining communicative performance from the perspective of the Speaker; contrast SWRT, which focuses on the relevance-theoretic *comprehension* procedure without offering a relevance-theoretic *production* procedure. Second, Straßheim imported Schutz's conception of typification into relevance theory, explaining how it is that a Speaker anticipates what will be relevant to the Hearer. Third, he argued for extending the types of experiences of Speaker and Hearer that are accounted for on the effect-side of the effort/effect ratio; he explained that "goal-related aspects of relevance" are essential to the ability of a relevance theory to explain communicative phenomena (Straßheim, 2010, p. 1436). Straßheim (2010, p. 1427-30) also made an effort to redefine "relevance," moving away from the effort/effect ratio to a ratio of continuation/contextualization. But as I have not adopted this alteration into CPR theory, I will not explore it here.

Though he did not identify it as one of his major contributions to relevance theory, I propose that Straßheim's first and most important contribution was to focus considerable attention on the Speaker's production of optimally relevant utterances, rather than only on the Hearer's interpretations. This marks an important difference from SWRT, which consistently focuses on the Hearer's interpretive processes but neglects the processes of production. Straßheim (2010) wrote: "Since the intended experience is highly selective, the *communicator* should, in her own interest, attempt to *suit her communicative means and ends to her addressee's dynamics of selections*, i.e. she should somehow diagnose and prognosticate what is relevant to her addressee under which circumstances" (p. 1419; emphasis mine). Unfortunately, Straßheim did not offer a relevance-theoretic explanation of how the Speaker produces her utterance.

Straßheim's second useful addition to relevance theory was an understanding of how a Speaker could predict what the Hearer will find to be relevant. In his view, "a relevance theory of communication needs to say more [than SWRT does] about how people know what would be relevant to others" (p. 1422). This is true given that we cannot have entirely congruent experiences.[8] As a solution, Straßheim (2010, p. 1423) proposed Schutz's theory of typification, an idealization that there is "a congruency of the systems of relevances" in Speaker and Hearer, summed up by this quotation from Schutz:

> Until counterevidence [sic] I take it for granted—and assume my fellow-man does the same—that the differences in perspectives originating in our unique biographical situations are irrelevant for the purpose at hand of either of us. (Quoted in Straßheim (2010, p. 1423).)

The idealization comes with a recognition that the counter-evidence Schutz mentioned

[8] And if we did, as Straßheim noted, there would be no need for communication.

could come at any time, and that if a Speaker or Hearer fails to recognize the counter-evidence, a miscommunication is possible or even likely. Schutz and Straßheim noted that Speakers and Hearers in most cases must immediately abandon the idealized similarity between them, but only to the extent that the evidence of their context and communications demands it.

A corollary to the use of typification arises from the fact that Schutz emphasized "connections ingrained through repetition and automatization, . . . relatedness framed by routines or habits, [and] the continuation of what went on before" in the construction of relevance (Straßheim, 2010, p. 1426). In other words, Speaker and Hearer will rely heavily on habitual performances and interpretations, barring counter-evidence that suggests they should vary from habitual practices. This is consistent with the "fluency heuristic" described by Gigerenzer and Brighton (2009). But if taken by itself, it suggests that the participants in my study would probably choose to produce habitually gendered stylistic performances in my sample, which they did not.

Straßheim's third useful addition to relevance theory was to argue that SWRT's definition of "cognitive effects" as alterations in the Hearer's store of assumptions was too "constrained to an inferential logic of assumptions." He saw that "Sperber and Wilson strictly define the 'effect' side of relevance (which shapes the whole notion) in terms of either deducing new assumptions, modifying the strength of old ones, or erasing old ones" (Straßheim, 2010, p. 1432). Recall from page 32 that "assumptions" are representations about the world that can be stated in propositional form. Straßheim claimed that such a limitation results in a theory that does not account for enough of what goes on in communication and contrasted it with the theory of Schutz, which considered "practical planning and acting" and emotions as important inputs to the relevance assessment (Straßheim, 2010, p. 1432).

As an example of such a non-assumption input, Straßheim made an extended argument that a relevance theory must account for "goal-related aspects of relevance," which he defined as:

> aspects of an explanation as to why a certain selective experience, such as some assumption or other, arose in an individual, namely those aspects of the explanation which refer the selection to some of the individual's goals. I use 'goal' in the widest sense. Both goals and assumptions are selective

> experiences, but they differ in a way which is broad and unproblematic: With an assumption, I *represent* or describe a certain state of affairs as (certainly, probably, possibly) true, whereas with a goal, I *plan* or *wish* for a certain state of affairs to remain or become true. . . . *Intuitively, an assumption seems to spring to my mind more easily when I see it as related to projects, intents or wishes that I happen to have at the moment. Conversely, when some conceivable assumption does not relate at all to a goal which is now strongly on my mind, I am quite unlikely to think of it* (Straßheim, 2010, p. 1433; emphasis mine).

As we saw in Section 2.4.1, SWRT—the relevance theory of Wilson and Sperber—addresses some of the concerns about classical pragmatics raised in Section 2.3: it dispenses with the Cooperative Principle of Grice as unnecessary; describes decoding and inference as parallel process operating online; and accounts for figurative language as a natural part of language comprehension. As we have seen in this section, SSRT—the relevance theory incorporating insights from Schutz and Straßheim—raises the issue of the Speaker's production of utterances and begins to explain how it is that the Speaker constructs her utterance so as to be relevant to the Hearer; and it also begins to address concerns that implicatures in classical pragmatics are propositional in form, taking special pains to explain the role of Speaker and Hearer's goals in communicative exchanges.

But while SSRT moves away from SWRT's heavy reliance on rational inference, SSRT does not offer a relevance-theoretic production procedure or explain how the heuristics discussed in Section 2.3.3 can influence production and comprehension as a result of the use of figurative language and other stylistic choices. As the study in this dissertation examines the utterances produced by the study participants, the theory accounting for their activity must have a production procedure.

Though the adaptations of SSRT offer considerable improvements to SWRT, they do not address all the concerns identified above. Consequently, I will offer a framework for cognitive pragmatic rhetorical (or CPR) theory in Section 2.5.

2.5 Cognitive pragmatic rhetorical theory

The goals of cognitive pragmatic rhetorical (CPR) theory are to offer a framework for describing human communicative activity in a way that can effectively be explored empirically, taking into account the metatheoretical and disciplinary commitments I described in Section 2.2. As we saw in Section 2.3, classical pragmatics is not quite up to the task. Essentially, CPR theory seeks to build a cognitive bridge between rhetoric and pragmatics. Section 2.4 described two models of relevance-theoretic pragmatics that have provided the foundations and main cables of such a bridge—the relevance theory of Sperber and Wilson (SWRT) and what I have called Schutz-Straßheim relevance theory (SSRT). What remains is to deck the bridge and open it for traffic. This section takes on that challenge.

CPR finally addresses all the concerns I have previously raised with antecedent theories: CPR theory takes account of a Writer's goals and beliefs about the world and her habitual practices and explains the efforts the Writer employs in finding, discovering, or inventing her communicative performances with the principle of relevance, which holds that she will expend effort in her writing choices that is commensurate with the accessibility and strength of the goals for which she writes and the effect she expects her writing to have on the cognition of the Readers(s). In the context of the study in this dissertation, CPR theory can explain the participants' communicative practices, how and why it is that they probably abandon habitual, gendered, communicative practices to adopt a style suited to the professional genre in which they write.

In brief, the CPR-theoretic view is that Speakers and Writers and Hearers and Readers are governed by the cognitive principle of relevance. I have described relevance as the ratio of positive cognitive effects and agent expects to the cognitive efforts in which the agent must engage to derive those effects. Here, I use "ratio" in the sense of a numerical fraction as a metaphor: I think of cognitive effects as the numerator in the relevance fraction and cognitive effort as the denominator of the fraction. Increasing cognitive effects increases the value of the fraction; increasing cognitive effort decreases the value of the fraction.

Section 2.5.1 first considers the components of the Speaker and Hearer's cognitive environments, a concept adapted from SWRT. The cognitive environment is essentially the agent's cognitive state at the moment of production or interpretation of an utterance, including his or her accessible assumptions and goals and his or her emotional state. CPR theory offers an alternative definition of "assumption" from that proposed by SWRT, and it extends the cognitive environment to include goals. That section also explains what CPR theory means by "accessible" assumptions and goals and recasts the *effort* variable from SWRT as "search costs"—the effort that an agent makes to search for productive or interpretive resources that are not presently accessible in his or her cognitive environment. The greater these costs, the greater the effort in the effects/effort ratio that defines relevance. Section 2.5.1 also accounts for the place that emotional states have in the cognitive environment. Finally that section describes the "imputed cognitive environment" as an alternative to SWRT's "mutually manifest cognitive environment." CPR theory's additions to and modifications of previous conceptions of cognitive environment serve to fill gaps identified above.

Section 2.5.2 then introduces the CPR-theoretic production and interpretation processes, which will be described in more detail in Section 6.2.

2.5.1 Components of cognitive environments

A central component of an agent's communicative production and interpretation is the agent's "cognitive environment," which I define this way:

(24) *Cognitive environment (CPR):* For an individual agent at a given time, the union of the set of assumptions and goals accessible to the agent and his or her emotional state.

I define each of these components and explain why it is included in this definition below. This definition of "cognitive environment" differs from Sperber and Wilson's. In SWRT, the cognitive environment includes only assumptions, but I reject both SWRT's limited definition of "assumption" and its restriction of the cognitive environment only to assumptions. After addressing those issues, I'll take up goals and explain what I mean by "accessible." Finally, I'll consider the issue of emotional states and imputed accessible cognitive environments.

Assumptions

Recall that for SWRT, assumptions are beliefs with propositional content. Sperber and Wilson (1995) define "assumptions" as "thoughts treated by the individual as representations of the actual world (as opposed to fictions, desires, or representations of representations)" (Sperber & Wilson, 1995, p. 2).

Representation itself is a problematic concept, one which Sperber and Wilson use frequently without really pinning it down. (See, for example, Sperber and Wilson (1995, p. 131).) Some cognitive scientists have challenged the utility of a representational theory of cognition. (See the discussion in the footnote at page 23.) For the moment, though, I'm prepared to accept the representational model of cognition in CPR, with the caveat that it could be replaced by a different model, if one presents itself. In any event, Sperber and Wilson conclude that assumptions can be presented in propositional form. For example, the moment after Mary walks into the office and says "Good morning!" to me, my cognitive environment (per SWRT) might include the assumption in (25):

(25) 'Mary just said, "Good morning!"'.

Note that if I entertain the thought in (26) or (27), I am entertaining not an assumption but a representation of a representation.

(26) 'I now believe that Mary just said, "Good morning!"'

(27) 'Mary wants me to believe that she wants me to have a good morning.'

But for Sperber and Wilson, assumptions are representations that describe the world as the agent believes it is, and not hypothetical circumstances or the state of mind of the agent or of another person. Thus, for them, (26) and (27) are not assumptions. Nevertheless, these hypothetical and meta-representations are the inputs and outputs of communication, and I will therefore not exclude them. In CPR theory, then, assumption is defined as follows:

(28) *Assumptions (CPR):* Thoughts that can be expressed in propositional form and are treated by the individual agent as representations of the world, including

the states of mind of the agent or others (meta-representations) and including hypothetical propositions.

CPR theory recognizes at least two attributes of assumptions that are important in their application: An assumption can be more or less accessible to the agent (see the discussion of accessibility below). And an agent can be more or less committed to an assumption; in other words, the degree of conviction she has about the truth of the assumption can vary.

Goals

Having addressed assumptions, we can turn now to goals. Goals are not a feature of the cognitive environment in SWRT, but Straßheim (2010) made an important contribution by noting that the calculation of "relevance" is difficult or impossible without accounting for the goals of the agent. Though Straßheim (2010) offered a definition of what he called "goal-related aspects of relevance" (see page 38), he did not offer a definition of "goal." This chapter is no place to offer a complete theory of goals for human action, but it seems necessary to offer at least a tentative definition that can be subjected to future refinement.

As a starting point, consider a definition from the realm of marketing psychology, which takes among its objects of study the means by which communications persuade consumers to make buying decisions:

> Goals are pleasant consequences (end states or otherwise) to be desired or unpleasant consequences to be avoided (negative goals). Goals are organized in hierarchies to facilitate their accomplishment. Lower-level goals (subgoals) are subordinate to higher-level goals. This means that satisfying lower-level goals helps in achieving higher-level goals. Higher-level goals represent the deep layer of... motivation (Gutman, 1997, p. 547, internal citations omitted).

As a preliminary definition of "goal" for CPR theoretical purposes, I offer (29):

(29) *Goals (CPR):* Consequences (end states or otherwise) desired or unwanted by an agent and capable of motivating an agent to action. [9]

As I will show in examples below, goals can be more or less accessible just as assumptions can; that is, the agent may or may not have a particular goal "in mind" during a particular communicative interaction. (See the further discussion of accessibility below.) Like assumptions, goals are also subject to varying degrees of commitment. An agent may have a some goals to which she is deeply committed and for which she will expend considerable energy. She may have other goals to which she is only loosely committed. The accessibility and commitment associated with goals governs the effect they have on relevance.

Accessibility and search costs

Goals that are accessible to an agent influence the way that she processes information (Peterman, 1997). But what does it mean to say that an assumption or goal is "accessible"? For a start, it means to acknowledge that it is possible for an agent to have assumptions and goals that are not influencing her present cognition to the full extent possible. And as we are taking SWRT as a starting point for this inquiry into relevance, it may be worth comparing the CPR theoretic "accessible" with SWRT's "manifest." For SWRT, an assumption (and only an assumption) is manifest to an agent at some time where "he is capable at that time of representing it mentally and accepting its representation as true or probably true," and "the environment provides sufficient evidence for its adoption" (Sperber & Wilson, 1995, p. 39).

Accessibility in the CPR theoretic sense instead captures the idea that not all assumptions and goals of an agent contemporaneously have the same effect on communicative practice. Consider the following examples. First, assumptions: Note that you may previously have adopted an assumption that is not accessible in your cognitive environment now. For example, you had probably adopted the assumption represented by (30) before reading (30) here:

[9] Future study of goals in this context probably needs to explore the way that goals—what we want or don't want—are shaped by values—why we want or don't want those outcomes (Gutman, 1997).

(30) In a right triangle, the sum of the squares of the lengths of the arms of the right angle is equal to the square of the length of the hypotenuse.

It is unlikely, however, that this assumption was as accessible in your cognitive environment before you read sentence (30) as was the assumption that you are reading this dissertation. By uttering (well, writing, actually) sentence (30) here, I have probably not added to the store of assumptions at your disposal. You could likely have produced a representation of the rule in sentence (30) in response to the need to calculate the distance from one corner of a room to the opposite corner using this rule of geometry. But by uttering sentence (30), I have prompted you to make that assumption accessible to you *now*. What's more, because of the associative nature of human memory, I believe I have changed the ease with which you will make related assumptions accessible in your cognitive environment. This is consistent with research in "semantic priming" (see Hutchison et al. (2013) for an overview of this literature). For example, I could expect you to be better prepared to comprehend a joke playing on the name of Pythagoras or some other allusion to geometry, if I made one.

Contrast the notion here of accessibility of an assumption with an agent's commitment to it. For example, the assumption in (27) may be one to which I'm only tentatively committed on the basis of the fact that Mary has said "Good morning!" to me. I may conclude that she could have said "Good morning!" out of polite indifference or even in spite of active animosity toward me. Thus, the assumptions in (26) and (27) may be equally *accessible* to me at the same time that I am *committed* more to (26) than to (27).

Second, consider goals: My goal of losing 20 pounds in the next two years would probably be relatively inaccessible in my cognitive environment while speaking to a colleague at the university about a grading dispute with a student. Similarly, my goal of achieving tenure may not be accessible to me while I am working with you to make revisions to an article we are writing to satisfy peer reviewers. Note that in the second instance, the goal of publishing our article may be a subgoal for my higher-level goal of achieving tenure, though it might certainly also be a subgoal to my higher-level goal of feeling a sense of accomplishment in my research. You might make either of those higher-level goals more accessible by referring or alluding to them in our conversation.

Here, too, commitment plays a role. I may be more committed to losing that weight than to obtaining tenure, or vice versa, and the relative level of commitment I feel toward these goals will affect the relevance of utterances related to them

Accessibility in the CPR theoretic sense, then, will always be a measurement of degree. And the easiest way to conceive of accessibility is to think of it in terms of "search costs": A goal or assumption is more accessible where the agent needs to search less in memory for it or to engage in less inferential reasoning to derive it. For many readers, exposure to (30) will "prime" assumptions related to Pythagoras and geometry. This conclusion is suggested by the semantic priming literature noted above (Hutchison et al., 2013). For most agents, mentioning their goals will prime them, I believe, making them more available to influence information processing and decision-making.

At this point, I want to return to the concept of the *heuristic*, discussed at some length in Section 2.3.3. There, we considered the heuristic as a "fast and frugal" means of reaching a conclusion, one that operates more or less automatically in human cognition and which might or might not be justified on rational grounds. This type of cognitive heuristic has sometimes been called a cognitive bias. The senses of the word "heuristic" include "[d]esignating or relating to decision making that is performed through intuition or common sense"—a sense consistent with that discussed above—but also the sense "[o]f, relating to, or enabling discovery or problem-solving" (Oxford English Dictionary, n.d.). The word's etymology goes to the Greek *euriskein*, "to find" (Oxford English Dictionary, n.d.). This is the same root as for the term "*heuresis*" the Greek name for the first rhetorical canon, also called "*inventio*" or "invention." I equate search costs with the costs of finding an assumption in memory or through inference. The cognitive heuristics discussed in Section 2.3.3 refer to low-cost methods for finding assumptions. But SSRT also identifies another source for low-cost searching: typification, those "connections ingrained through repetition and automatization,. . . relatedness framed by routines or habits, [and] the continuation of what went on before" (Straßheim, 2010, p. 1426). (See Section 2.4.2.) This leads (finally) to a tentative definition of "accessible" in CPR theory:

(31) *Accessible (CPR):* An assumption or goal is more accessible than another if it is available to influence information processing and decision-making with lower

search costs than the other assumption or goal; in general, search costs are lower where a cognitive heuristic (a la Gigerenzer and Brighton (2009)) or the agent's habitual practice leads to the assumption or goal, or the assumption or goal is closely related to one already made accessible by previous communication.

Consider this example: I described above a hypothetical situation where my goal of losing 20 pounds in the next two years would probably be relatively inaccessible in my cognitive environment while speaking to a colleague at the university about a grading dispute with a student. Perhaps that colleague wishes to point out the amount of work in which I will have to engage to sort the matter out by uttering (32).

(32) You'll be running back and forth across campus to get this sorted out with the registrar's office!

Imagine, though, that the colleague, who knows me well, is trying to lighten the mood of the conversation. Instead of (32), she might utter (33).

(33) Well, all the running around you'll do to sort this out may actually help with that weight you are concerned about losing.

Interpreting (32) would come with less search cost for me than interpreting (33), because my weight-loss goal was not previously as accessible to me in this conversation as was the need to go across campus to resolve the grading problem. However, once my colleague has uttered (33), my weigh-loss goal and the associated assumptions and subgoals become more accessible, and in fact, we might shift the whole topic of our discussion to these goals, leaving behind the topic of the grading dispute.

Emotional states

Rhetoricians since Aristotle have overtly recognized the role of emotion in communication and persuasion. Among his three artistic proofs was *pathos*, or appeal to emotion (Aristotle, 2007). Dascal and Gross (1999) suggested, rightly I think, that rhetoric and pragmatics need to account for emotion in the production and interpretation of communicative performances. They limited that assertion, though, to claiming that the emotional aspects of communication should be subject to rational inference. I believe it's important to distinguish the conception of *emotional states* and their effects on the production and interpretation of communication from assumptions about emotional states, which can be the objects of rational inference.[10]

Consider this example: Waleed has had a contractor do work on his home. The contractor sent him a bill for more than three times the estimate provided before work began, which is more than he can afford to pay, and he has contacted the owner of the contractor to object. She is immovable on the amount of the bill, and finally issues the threat in (34):

(34) Either you pay this bill in full, or I will have my lawyer file a mechanic's lien against your home and foreclose in court, and then you'll be liable for my attorney fees, too.

At this point, Waleed might *feel* intensely threatened, angry, and frightened of losing his home and being subject to even more expense. On the other hand, if he has reason to believe the contractor's threats are hollow (that she has no such legal rights), he might not have those feelings at all. Thus he may (or may not) adopt any of the following assumptions:

(35) 'I feel frightened.'

(36) 'The contractor's threat has made me feel frightened.'

(37) 'The contractor has made her threat to make me feel frightened.'

These assumptions can be the objects of rational inference. For example, once Waleed adopts (35) and (36), he might rationally infer (37). The feelings Waleed is *experiencing* can affect his cognition in a wide variety of ways. For example, fear may cause him to avoid the threat-condition and agree to pay the contractor. But note that Waleed may experience the feeling of fear without adopting any of the assumptions in (35) through (37), and that he might adopt the assumption (37) without actually feeling frightened.

[10] Note that I have not generally taken up the issue of emotional states in this dissertation. While I believe a discussion of it is essential for CPR theory to be a general theory of text production and interpretation, I will leave that matter to Section 7.4 and future studies.

The effect of the Speaker's utterance on the emotions the Hearer experiences is therefore distinct from the assumptions that the Hearer has about his emotions. CPR theory maintains this distinction by evaluating emotions as cognitive effects in their own right, separately from any propositions about them.

Imputed cognitive environments

An agent interacting with another person adopts assumptions about what the other's cognitive environment includes. In other words, a Speaker has expectations about what assumptions and goals are accessible to the Hearer and what emotions he is experiencing. These assumptions about the Hearer allow the Speaker to construct utterances that she can expect the Hearer will find relevant and that she can expect to influence the Hearer's cognitive environment. Conversely, the Hearer's assumptions about the Speaker's cognitive environment are accessible to him and assist him in interpreting the Speaker's utterances. These assumptions by an agent about the cognitive environment of another agent are called "imputed cognitive environments" in CPR theory (see (38)).

(38) *Imputed cognitive environment (CPR):* The assumptions that an agent has about the cognitive environment of another agent or group of agents, including the agent's assumptions about the cognitive environment that is being imputed to her. In other words, these are the assumptions that the agent believes she and other the other agents all share about their current situation

This concept is modeled after the "mutually manifest" cognitive environment of SWRT, but I have avoided the terminology from SWRT. The name "mutual manifestation" suggests that the two agents actually share assumptions and that they know they share them. Sperber and Wilson (1995) take pains to clarify that this is not the case: What is mutually manifest between Speaker and Hearer may be different for Speaker and Hearer. In other words, what the Speaker thinks the Speaker and Hearer both believe may be different than what the Hearer thinks they both believe. Thus I believe the "mutually manifest" name is unnecessarily confusing and have chosen "imputed" cognitive environment instead. This conception of a mutually shared environment imagined by one party to it is consistent with the conception of audience in Perelman and Olbrechts-Tyteca (1969): They note that for a speaker, her audience is always a construction of her own, based on what she knows (or thinks she knows) about the audience.

An agent may construct an imputed cognitive environment with relation to another agent (or group of them) based upon a variety of assumptions and stimuli. For example, if two agents are interacting in a typified situation (a la SSRT), each can be expected to understand the underlying assumptions held by the other. But as Schutz noted, the two agents cannot have had the same experiences of even these typified environments, so their imputed cognitive environments cannot be entirely congruent.

2.5.2 CPR-theoretic production and interpretation

Under CPR theory, I argue that the production and interpretation of utterances is governed by relevance, but I propose a slight alteration to the formulation of the concept of relevance. As with SWRT, relevance in CPR theory is a metaphorical ratio of effect to effort. But the workings of this ratio are slightly different for the Speaker than for the Hearer. This extension of the previous relevance-theoretic approaches is important for my study, because I must ultimately explain the choices that study participants here made when producing texts that they believed would stylistically conform to the legal genre they were attempting to write. This section introduces the CPR-theoretic production and interpretation processes. They are described in more detail in Section 6.2, just before I apply those concepts to explain *gender*, *genre*, and the findings in this study.

A Speaker seeks by her utterance to change the cognitive environment of the Hearer. The changes the Speaker seeks to work on the Hearer's cognitive environment will depend on which of the Speaker's goals are most accessible to her when she is speaking and the strength of her commitment to them. The Speaker's cognition is governed by the principle of relevance. She seeks to maximize the ratio of effect to effort. So I offer the CPR-theoretic production procedure:

(39) *CPR-theoretic production procedure:* The Speaker should incur search costs in producing her utterance proportional to the effect she expects it to have on the Hearer's cognitive environment, taking into account the imputed cognitive

environment, the Hearer's likely assessment of the utterance's relevance to the Hearer, and the accessibility and strength of the goal the Speaker is attempting to advance.

See Section 6.2 for elaboration and examples. Note that this procedure presumes that the Speaker can assess at the outset what level of search costs she should invest in producing the utterance; she then produces the utterance.

Relevance for the Hearer consists of the ratio of effects to efforts, too. But unlike SWRT, CPR theory does not hold that a Hearer's assessment of cognitive effects is limited to the enhancement of the assumptions available to the Hearer. On the contrary, CPR theory acknowledges that the Hearer may be engaged in the communicative exchange to obtain something more or other than information.

CPR theory also measures the Hearer's effort in terms of search costs: The greater the extent to which he can identify the Speaker's utterance as an example of a habitual utterance pattern, the lower the effort required to interpret her utterance.

However, because the Hearer can presume that the Speaker has attempted to make her ostensive communication optimally relevant to him, he can continue generally to rely on the SWRT comprehension procedure, which directs him to follow a path of least effort in interpreting the utterance and to stop when his expectations of relevance are met. (See the discussion regarding (15) and (16), above.) The CPR theoretic comprehension procedure accounts for the accessibility and commitment of the Hearer's goals, though, and can thus be defined this way:

(40) *CPR theoretic comprehension procedure:* The Hearer should follow the path of least effort in interpreting the Speaker's utterance, testing interpretative hypotheses in order of accessibility, stopping when his expectations of relevance are satisfied, taking into account the imputed cognitive environment and the goals the Hearer is attempting to advance.

See Section 6.2 for elaboration and examples.

2.6 Conclusion

This chapter has introduced cognitive pragmatic rhetorical (CPR) theory, which provides a framework for analysis not available in classical pragmatics, in the relevance theory of Sperber and Wilson (SWRT), or in the extension of that theory by Straßheim (SSRT). CPR theory takes account of the Writer's cognitive environment: her accessible goals and assumptions (propositional beliefs about the world) and the extent to which she is committed to them; her emotional state; the cognitive environment(s) she imputes to her Reader(s), including any expectations the Writer thinks the Reader has about the Writer's performance; and the Writer's express and tacit knowledge about likely cognitive effects of her performance on the Reader(s).

This chapter has grounded CPR theory in three different disciplines: pragmatics, rhetoric, and cognitive science. But principally, I see CPR theory as a bridge between rhetoric and linguistic pragmatics, with SWRT as the foundations and main cables and SSRT as the suspension cables. CPR theory completes this bridge. The first traffic on it is the empirical study in this dissertation.

CPR theory explains the findings in the study in this dissertation. The students who participated in the study were completing a year of training in the legal profession, during which they had received exposure to some legal genres and the linguistic register of legal writing; they likely had some assumptions (beliefs) about the genre and register in which they were expected to write. We presume that they came to their law studies with habitual, gendered, communicative practices.[11] Varying from these habits would normally reduce the relevance of the task in which they were engaged by increasing the effort without a commensurate increase in effect. But here, the writing assignments come at a critical juncture and come with high stakes. These facts probably change the relevance ratio: Where there is a greater effect to be achieved—namely success on a project at the heart of a goal that is accessible to the student and to which she is firmly committed—greater effort is warranted—namely, attention to the linguistic register of the law and suppression of habitual practices.

Before discussing the design and findings of the study in this dissertation in some detail in Chapters 4 and 5, I review some of the literature studying gender differences in

[11] See Section 3.7 for the tentative conclusion that men and women write differently in informal contexts and particularly the folk perception that men and women communicate differently.

language in Chapter 3. After presenting the findings, I analyze them using CPR theory in Chapter 6. Finally, in Chapter 7, I consider implications of the empirical study in this dissertation and of CPR theory more broadly.

Chapter 3

Gender differences in writers' choices

3.1 Introduction

This dissertation applies cognitive pragmatic rhetorical (CPR) theory in an effort to explain the absence of patterns of gender difference in texts produced by students in a pre-professional training context. Such patterns have been observed by researchers in previous studies. As we shall see below, we can conclude tentatively that men and women operating in typical social contexts engage in communicative activities that vary with their genders. But the participants in this study did not exhibit such patterns of difference.

CPR theory takes account of a Writer's goals and beliefs about the world and explains the efforts the Writer employs in finding, discovering, or inventing her communicative performances with the principle of relevance, which holds that she will expend effort in her writing choices that is commensurate with the accessibility and strength of the goals for which she writes and the effect she expects her writing to have on the cognition of the Readers(s). In other words, relevance is a ratio of desired cognitive effects to cognitive effort. In this study, CPR theory explains that the goals of the Writers, particularly their desire to achieve success in their first year of professional training; and CPR theory explains their exacting attention to the conventions of legal writing and their abandonment of habitual gendered communicative practices.

This chapter considers the wisdom of doing such an empirical study and also the research constructs essential to performing it, taking into account methodological limitations of many previous studies.

A preliminary question is important: Should we do studies of gender difference in communication? Some have argued that such studies take a sex or gender binary as a presumption and essentialize differences between the genders (or sexes) (DeFrancisco, Palczewski, & McGeough, 2014). But others argue that quantitative studies of variation provide important epistemic purchase (Hultgren, 2008). Section 3.2 takes up this debate. I conclude that the arguments against such studies are outweighed by the potential advantages arising from well-constructed studies.

Section 3.3 describes the Argamon/Koppel 02/03 study (Argamon, Koppel, Fine, and Shimoni (2003), Koppel et al. (2002)),[1] which provided the motivation for the empirical study in this dissertation and also served as a cited example for many later studies. The Argamon/Koppel 02/03 study is described in considerable detail, including an overview of machine learning algorithms, an analytical tool used in Koppel et al. (2002), in many later studies, and in the study in this dissertation. That section also highlights the limitations of the Argamon/Koppel 02/03 study, which are common to many other studies: First, the authors failed to ground the gender categories they used theoretically and to explain how gender categories were ascribed to the texts they studied. Second, though they recognized that text genre was a confounding variable in their study of gender differences, they failed to seize the opportunity to explore and explain its role. And finally, they failed to account for the fact that the published texts they studied might not be authored by individuals or that they might be heavily edited by persons unidentified in the corpus studied and that the editors and co-authors might be of different genders than the authors who received the bylines.

The next three sections take up each of those limitations with regard to other studies and argue for the solutions adopted for the empirical study in this dissertation. Section 3.4 looks at limitations of other studies stemming from their failure to theorize or to

[1] Throughout this dissertation, Koppel et al. (2002) and Argamon, Koppel, Fine, and Shimoni (2003) and the underlying study are occasionally referred to as as the *Argamon/Koppel 02/03 study* because both essays relied up on the same data set prepared by these researchers. Note that Argamon, Koppel, Pennebaker, and Schler (2007) is also cited in this dissertation; this 2007 study, cited and discussed elsewhere in this dissertation, is a different study than the 2002/2003 one and not what is meant in references to the *Argamon/Koppel 02/03 study*.

explain how they ascribed *gender* as a category and offers my proposal for a framework for doing both things. The section reviews studies in technical and professional communication, studies that cited Argamon, Koppel, Fine, and Shimoni (2003) or Koppel et al. (2002) but did not employ its machine-learning methods, and studies that cited Koppel et al. (2002) and also used machine-learning. Section 3.5 examines the ways in which previous studies failed to account for the genre knowledge of text authors, contextual knowledge that CPR theory predicts could have a powerful effect on communicative performance. Finally, Section 3.6 provides a brief explanation of my solution for what I call the "single-author" problem.

Before embarking on a description of the study in this dissertation in Chapters 4 and 5, the reader can expect me to assess, given the studies described below, whether in general there are significant differences in the communicative practices of women and men. I conclude tentatively in Section 3.7 that men and women do, in general, make communicative choices that vary with their genders.

But the first question, as noted above, is whether it is wise even to do this sort of study. The next section takes up that question.

3.2 Should we do gender-difference studies?

DeFrancisco et al. (2014, p. 59) warned against the risks of doing studies that focus on "sex difference only" in communication research. First, they contended that researchers asking the sex-difference question are "likely asking the wrong question and will only reinforce essentialist views of gender/sex in communication" (p. 59). Second, they criticized researchers who conflate *sex* and *gender* as variables. Third, they claimed that studies looking for statistical differences between communicative practices of gendered persons may overlook the very substantial similarities in their communications. Finally, they asserted that studies focused on sex-difference are missing the possibility of an "instersectional analysis," would would "explore whether interdependent ingredients serve to influence the gender performance in unique ways yielding unique privileges [and] inequalities" (p. 60).

Their critique is thoughtful, but it overlooks important counterarguments. First, whether researchers are asking the right or wrong question probably depends on the context of the question. But to contend that sex-difference (or gender-difference) studies will "reinforce essentialist views" is to assume that such studies will find patterns of differences. If a well designed study shows, however, that there are few or no differences between the communications of female and male authors when the writing context is taken into account—exactly what the study in this dissertation does—that seems a powerful anti-essentialist argument about sex-differences. Showing that males and females are equally capable of adapting to the communicative conventions of a new professional identity is one way of reinforcing the view, which I believe DeFrancisco et al. (2014) embraced, that patterns of difference in communication arise in contexts where the participants are in incongruent power relationships.

Second, for reasons I explain below, I contend that researchers in communication are almost never studying *sex differences* but rather *gender differences*. Communications researchers are unlikely to engage in the kind of physical examination of study participants that would allow them to make an ascription of a sex category. At best, they can make educated guesses by identifying external physical characteristics—which can be misleading—or asking participants to identify their own sexes—which calls on the participants to make gender performances. That is, when the participant responds "male" to the question on a survey that asks his sex, he is providing evidence of what he wants the researcher to believe his sex is, which is to say, he is making a gender performance.

Third, the only way to show whether there are communicative similarities and differences between men and women in similar contexts is to look. And if there are *statistically* significant differences, one can then proceed to ask whether the statistically significant differences are actually "socially significant" (DeFrancisco et al., 2014, p. 60). The absence of well designed studies that look for gender differences has two possible consequences. If the studies would have shown folk beliefs about men's and women's communicative practices are wrong, they would have proved useful for the very project for which DeFrancisco et al. (2014) advocated. If, on the other hand, well designed studies of gender differences would discover differences, an intersectional theory of gender and communication should either account for them based on other "interdependent ingredients" or acknowledge that there are deeper differences between the genders (or sexes) than the theory previously admitted. In either case, an opportunity for knowledge making is lost.

Finally, if a study can show—as I believe the study in this dissertation does—that study participants need to have very little in common before their writing exhibits none of the patterns of difference seen in previous studies, researchers and educators are in a position to help students understand how they can alter their communicative performances (if they seek to do so) to achieve their goals. Of course, richer ethnographic studies of the kind for which DeFrancisco et al. (2014) advocate still have incredible value. Nevertheless, knowing in advance the ease with which one of those variables can cease to have an apparent effect on the communicative practices of participants in some contexts is itself valuable to the ethnographic researcher seeking to provide a thicker description of communicative practices from an intersectional perspective.

I adopt the views explained by Hultgren (2008), who called the strand of research in which the empirical study in this dissertation lies "correlational sociolinguistics." She defined the practice and also noted its infrequent use in recent years:

> [Correlational sociolinguistics is] research within the field of language and gender that (1) takes binary sex as a legitimate starting point for analysis, and (2) relies on quantification to identify general patterns of variation between male and female speakers. Whilst such methods are still prominent within variationist sociolinguistics, they have virtually been abandoned within language and gender research.

Hultgren argued that this approach to studying sex (or gender) and language should work alongside more ethnographic (and other) approaches. This approach, she said, offers three important advantages:

> First, it strives to reduce researcher interference in the data; secondly, its reliance on quantification means that it has the capacity not only to report on sex differences but also on the absence of such differences; lastly, it is not as guilty as 'gender in discourse' approaches of theorising in a void (Hultgren, 2008, p. 34).

The second of these advantages I discussed above. As for the first and third, they perhaps require further elaboration.

Hultgren's choice of the expression "researcher interference in the data" is unfortunate; I believe her argument on this point owes more to notions that researchers should be able to expose their methods of analysis and interpretation in such a way that other researchers can interrogate them and enter into conversation with them. As she said:

> In correlational sociolinguistics, it should be possible to justify why you have coded data in the way you have and, more importantly, to communicate this justification to a wider audience. In discourse analysis it is often not possible for readers to trace the analyst's moves from data to interpretation and hence much interpretation remains clandestine (Hultgren, 2008, p. 35).

The term "clandestine" also suggests to me a predisposition to *distrust* discourse analysis (and other qualitative approaches). My view is rather that such studies play an important role, but that they should enter into dialog with quantitative studies so that each can support the other: Quantitative studies can help on a larger scale to validate the findings of small qualitative studies; and qualitative studies can provide an understanding of subtle variables that can greatly affect the findings in quantitive research if they are not accounted for.[2]

The third advantage Hultgren (2008) identified for quantitative studies of this kind is "avoiding theorizing in a void." Here, she was largely noting that an academic view "that there is no prediscoursive gendered reality. . . fails to resonate with most people outside academia" (p. 37). I read her argument as saying that if the academy intends to deny the everyday conceptions of reality of folks outside academia, the academy needs to speak from a ground that the rest of society will respect. This means acknowledging common public conceptions of sex or gender difference. And in my view it means using the kind of data that the public accepts as authoritative and that has given rise to the essentializing gender-difference narrative that now prevails—using quantitative data is fighting fire with fire where the cool water of academic theory won't do.

Of course, a researcher doing quantitative gender-difference studies in communications should examine the studies of others for methodological soundness and should construct her own studies with methodological rigor. This is done less frequently than we might hope. Studies that are constructed carefully with regard to many aspects

[2] I don't mean to suggest an order or priority by discussing the value of quantitative studies before that of qualitative studies.

often fail to theorize gender or sex and often fail even to explain how those categories were ascribed to study participants or artifacts. The following sections take up those methodological limitations and others.

3.3 The Argamon/Koppel 02/03 study

Above I mentioned a study of Shlomo Argamon, Moshe Koppel, and their colleagues, which I have called the Argamon/Koppel 02/03 study (Argamon, Koppel, Fine, & Shimoni, 2003; Koppel et al., 2002). Koppel et al. (2002) reported a study on a subset of the British National Corpus (BNC) to investigate whether texts published in print (which the researchers referred to as "formal texts") could be automatically classified based on the author gender. They used a machine-learning algorithm called Winnow, trained on texts previously classified by gender and genre, to create a linear model of the attributes most likely to be useful in classifying previously unseen text instances by gender and genre. Argamon, Koppel, Fine, and Shimoni (2003) reported the results of a statistical study of the same texts, looking for lexical and quasi-syntactic features that varied with the author's gender. As the Argamon/Koppel 02/03 study was an inspiration for the study in this dissertation, and because many studies looking for gender differences in written communication since have cited to it, this section describes it in some detail: the machine-learning analysis of Koppel et al. (2002) in Section 3.3.1, the statistical analysis of Argamon, Koppel, Fine, and Shimoni (2003) in Section 3.3.2, and their limitations in Section 3.3.3.

Argamon, Koppel, Fine, and Shimoni (2003) and Koppel et al. (2002) both used attributes based on the style of texts, rather than their content. They used the relative frequency of each of 76 part-of-speech tags, 405 function words, and 600 part-of-speech trigrams and bigrams as attributes or variables for their study (Koppel et al., 2002, p. 404). Each text in the corpus was therefore represented by 1,081 attributes (Koppel et al., 2002, p. 404).

Each text from the BNC is broken into *tokens*. Types represent classes of which tokens are instances. Consider the sentence, "The red bottle goes in the red bin." It consists of nine tokens (including the period at the end) but only seven types ("the" and "red" are repeated) (Jurafsky & Martin, 2009, p. 86). Each token in a text thus represents one instance of a word or punctuation mark. Each token is then assigned a part-of-speech ("What is the BNC?," n.d.). The 61 parts of speech used in the BNC are those of the C5 tagset, which provides a much more granular list of parts of speech than readers may be accustomed to from sentence diagramming (Leech, n.d.).[3] So for example, there are more than 20 different tags for verbs, including "VVD The past tense form of lexical verbs (e.g. *forgot, sent, lived, returned*)"; "VVI The infinitive form of lexical verbs (e.g. *forget, send, live, return*)"; and "VDG the *-ing* form of the verb DO: *doing*" (Leech, n.d.). There are tags for attributive, comparative, and superlative adjectives, for punctuation marks, and for a few key function words, like *not*, and the infinitive marker *to*.[4]

Function words are a rather difficult category to define. Jurafsky and Martin (2009) defined them as tending "to be very short, occur frequently, and often have structuring uses in grammar" (p. 125). They commonly come from *closed classes*, parts of speech that "have relatively fixed membership," like pronouns and prepositions (p. 124-25). See Appendix C for the complete list of function words the researchers in the Argamon/Koppel 02/03 study used. Among them are many that seem quite obviously to fit into these categories:

- Articles, other determiners, and quantifiers, e.g., *a, an, billion, eight, eighteen, eighty, every, few, many, most, these, those.*
- Conjunctions, e.g., *and, because, but, if, nor, or.*
- Pronouns, e.g., *her, her, hers, himself, I, she, they, we, you.*
- Prepositions, e.g., *about, above, beneath, from, of, within.*[5]
- Forms of common functional verbs, e.g., *be, can, do, go, have, may.*

[3] As I noted in the previous paragraph, the Argamon/Koppel 02/03 study used 76 tags. The reason there are 15 more tags than there are parts of speech is that sometimes two parts of speech are assigned to a single token by use of a double or ambiguous tag. That results in 15 more tags than there are parts of speech.

[4] For the study in this dissertation, I have used the Penn Treebank tagset. The Penn Treebank tagset appears in Appendix B; my explanation regarding its use in this study appears in Section 4.3.2.

[5] One of the reviewers of this dissertation noted that there may be a distinction here between multi-syllable prepositions that play more of a lexical function and single-syllable prepositions that play more of a grammatical function. I have not yet considered the implications of such a claim.

- Common adverbs, particularly adverbs of time or expressing probability or frequency, e.g., *actually, again, certainly, doubtfully, ever, eventually, later, never, subsequently.*

- Interjections, e.g., *ah, ai.*

In contrast to function words are "open class" or "content" or "lexical" word types. These word types tend to convey more of the semantic content of a text. Some items on the function word list from the Argamon/Koppel 02/03 study seem less functional and more lexical:

- Adjectives, e.g., *actual, certain, definite, doubtful, possible, previous, rare.*

- Verbs, e.g., *bear, bring, come, enter, forgo, get, give, let.*

- Nouns, e.g., *billionth, example, instance, fact.*

Finally, part-of-speech bigrams and trigrams are a way of viewing quasi-syntactic data about the text. A POS bigram is a two-token window into the text, showing how frequently ordered pairs of parts of speech appear in the text. A trigram is a three-token window.

To put it all together, let's consider the processing of one sentence according to this model as an example: *My aunt's pen is on the table.* If this sentence appeared in the BNC, it would already be tokenized and tagged with parts of speech. See Figure 3.1.

The tokens *my, 's, is, on,* and *the* are instances of function words (highlighted in light purple in Figure 3.1). Each token, including the period, is assigned a POS tag: PNP for the personal (possessive) pronoun *my*, NN1 for the singular common nouns *aunt, pen,* and *table*, VBZ for the third-person singular present tense form of the verb *to be*, and so forth.[6] Finally, each sequence of three POS tags counts as a POS-trigram. Thus, the parts of speech of the first three tokens form a trigram, PNP_NN1_POS—*my aunt 's*—as do those of the second through fourth tokens, NN1_POS_NN1—*aunt 's pen.* The seven resulting trigrams are shown cascading down from the original sentence in Figure 3.1.

[6] Note that I have used the Penn Treebank part-of-speech tags in this example. See Appendix B for a complete list of these tags and their meanings.

	My	aunt	's	pen	is	on	the	table	.
	PNP	NN1	POS	NN1	VBZ	PRP	AT0	NN1	PUN
1	PNP	NN1	POS						
2		NN1	POS	NN1					
3			POS	NN1	VBZ				
4				NN1	VBZ	PRP			
5					VBZ	PRP	AT0		
6						PRP	AT0	NN1	
7							AT0	NN1	PUN

Figure 3.1: **Sample sentence tokenized with function words, parts of speech, and POS trigrams identified**

3.3.1 The Koppel et al. 2002 machine-learning study

What I have called the Argamon/Koppel 02/03 study consists in two articles, Koppel et al. (2002) and Argamon, Koppel, Fine, and Shimoni (2003), which the authors prepared based upon their analysis of the same corpus. The first of these to appear was Koppel et al. (2002), which used a machine-learning algorithm (MLA) and explored the question of whether an MLA could successfully classify previously unseen texts by author gender after the MLA was "trained" on text instances that had been classified by author gender. This study used the variables or "features" described in Section 3.3. This section first provides a basic overview of MLAs, which will be useful both for understanding Koppel et al. (2002) and as context for the study in this dissertation, which is described more fully in Chapter 4. It then presents the findings of Koppel et al. (2002). Section 3.3.2 then takes up the second study, Argamon, Koppel, Fine, and Shimoni (2003). Section 3.3.3 describes methodological limitations that affected both of those essays.

Introduction to machine learning

Koppel et al. (2002) used a method from computer science and artificial intelligence called a *machine learning algorithm* (MLA). MLAs are powerful tools for identifying patterns involving many features that can serve as a basis for distinguishing or categorizing texts. For purposes of Koppel et al. (2002) and the study described in this dissertation, machine learning is the use of computer algorithms to identify patterns—possibly very subtle ones—in texts to determine whether the texts can be classified based on lexical, quasi-syntactic, or other textual characteristics so that the classifications are consistent with a category variable, in this case author gender. Witten, Eibe, and Hall (2011) offered this description of machine learning:

> [Machine learning's] input takes the form of *concepts*, *instances*, and *attributes*. We call the thing that is to be learned a *concept description*. The idea of a concept, like the very idea of learning in the first place, is hard to pin down precisely, and we won't spend time philosophizing about just what it is and isn't. In a sense, what are trying to find—the result of the learning process—is a description of the concept that is *intelligible* in that it can be understood, discussed, and disputed, and *operational* in that it can be applied to actual examples (Witten et al., 2011, p. 39; emphasis in original).

The attributes are also sometimes called *features* and correspond to the common concept of *variables*. As with any investigation, the researcher must select the attributes or variables to be considered. The *attributes* for a study may include relative frequencies of certain lexical items (or *word types*) and quasi-syntactic characteristics. The *concepts* to be described in Koppel et al. (2002) were gender and text genre; the study in this dissertation explores the same concepts, though as we shall see below, this dissertation conceives of the research constructs *gender* and *genre* differently than Koppel et al. (2002) did. Because Koppel et al. (2002) explored a *classification* problem—whether texts could be classified according to author gender and text genre—the concepts of gender and genre were also the *classes* of the instances in their work. In this dissertation, gender is a class, but genre is not, as the texts in my study are all of the same genre.

Paper	Gender	Tri_DT VBN NN	Tri_POS NN TO	Tri_WDT MD VB
1001	1	0.005374899	0.00107498	0.00107498
1003	0	0.015328933	0.002554822	0.000851607
1004	0	0.009760065	0	0.001626678
1005	1	0.006852668	0.00783162	0.000978953
1007	1	0.037396122	0.002770083	0.001385042
1008	0	0.017526777	0.00194742	0.00097371
1009	1	0.023426061	0.001171303	0
1010	0	0.012990257	0.001998501	0
1011	1	0.00216041	0.00216041	0
1012	0	0.017033357	0.002838893	0.002838893
1013	1	0.030337079	0.001123596	0.004494382

Figure 3.2: **Simplified spreadsheet showing abstraction of texts before application of machine learning**

Note that the machine learning algorithm does not evaluate the unit of study directly. Rather, it evaluates an abstraction of that object, a data set consisting of attribute labels and a value for each attribute for a given instance. The most accessible visual representation of this approach is a spreadsheet, where the rows represent instances and the columns represent attributes. Each cell represents the value of an attribute for an instance. The researcher must select the attributes for which values are measured. A much simplified example using texts from this study appears in Figure 3.2. In this case, the column labeled "Paper" records a unique ID number for each text, which represents an *instance*; the "Gender" column indicates the *class label* for the instance; and the three *features* or *attributes* of each instance are recorded in the columns labeled "Tri_DT VBN NN," "Tri_POS NN TO," and "Tri_WDT MD VB." (The meanings of these feature labels are unimportant for the moment.)

The outputs of machine learning algorithms are concept descriptions, what Witten et al. (2011) also call "knowledge representations" and describe as "descriptions of the structural patterns in the data"(Witten et al., 2011, p. 61). The form of the concept description depends on the input data and the machine learning method applied to it.

Witten et al. (2011) provided accounts of several types of machine learning algorithm. Here are brief summaries of their methods and resulting knowledge representations:

- A *linear model* uses the training set to generate a knowledge representation in the form of a linear regression equation. In a classification task, each attribute is evaluated based on the training set data for the weight it should be given in an equation that will define the decision boundary between the classes.

- A *Bayesian probabilistic model* is a set of conditional probabilities associated with classes given the priors and probabilities of the priors given the class labels. The concept description consists of a set of variables, arcs between them, and probabilities on each of the arcs.

- A *decision tree* uses the training set to generate a model in which an instance in the test data is evaluated according to a sort of flow chart, each node of which tests a single attribute of the instance. The knowledge representation is the flow chart.

- A *rule-based* model uses the training set data to generate a series of *if-then* rules. Unlike the attribute test in a decision tree, the antecedent of a rule can refer to multiple attributes. For example, *If Attribute1 <1 or Attribute3 >5, then the instance is an X.* A decision-tree model can easily be converted into a rule model, but because of this possibility of disjunction in a rule antecedent, rule models cannot necessarily be converted into decision tree models easily.

- An *instance-based* method does not build a model when exposed to the training set. Instead, when exposed to an instance from the test set, it calculates which instance in the training set is most similar to the test instance and classifies the test instance the same as that training instance. The similarity is based on a calculation of *mathematical distance* between instances. Thus, in a sense the knowledge representation is the set of the previous instances.

- Hybrid methods may combine the foregoing. For example, *support vector machines* and *locally weighted linear learning* are hybrids of linear and instance-based methods.

Once Koppel et al. (2002) had preprocessed the texts as described above and represented them as a collection of values for the attributes of each instance, they used the Winnow algorithm to examine each attribute in each text in a set of training texts where Winnow was given the gender and genre; it assessed the utility of each feature in an iterative process, eventually assigning a weight to each feature's value in determining whether a text is by an author of one gender or the other or is of one genre or the other (Koppel et al., 2002, p. 405).

The Koppel et al. 2002 study findings

Winnow categorized a corpus of more than 500 texts, averaging about 34,000 tokens per text, by the gender of their authors with observed agreement of 77.3%.[7] Koppel et al. (2002) noted that differences between fiction and non-fiction were "generally greater than the difference between male and female writing styles and thus training on fiction and non-fiction documents together actually harms results" (Koppel et al., 2002, p. 406). When Winnow trained on non-fiction documents only, its accuracy of categorizing non-fiction texts by gender increased to 82.6%; when trained on fiction only, its accuracy of categorizing fiction texts by gender increased to 79.5%. Koppel et al. (2002) found "that the male indicators are largely noun specifiers (determiners, numbers, modifiers) whereas the female indicators are mostly negation, pronouns, and certain prepositions."

Koppel et al. (2002) noted the difference in performance when attempting to categorize the texts by genre instead of gender:

> An interesting phenomenon... is that the differences between male and female usages of various features parallel more extreme differences between fiction and non-fiction: determiners, which are used more by men, are used more by all authors in non-fiction; pronouns and negation, which are used

[7] Note that the authors of this study did not say exactly what the implications of this level of observed agreement were. In my view, machine learning is often used to achieve institutional or organizational objectives. So for example, a company might want to identify the genders of folks writing comments on its web site so that it can market particular products to them. In that instance, the company's application and the costs of its efforts would take into account this level of observed agreement. From a theoretical or knowledge-building standpoint, there is no clear implication of this level of observed agreement. But see the discussion in Section 4.4.2 of the threshold of practical significance selected for the empirical study in this dissertation.

> more by women, are used more by all authors in fiction. The extreme differences between fiction and non-fiction suggest that distinguishing between the two genres ought to be an easier task than distinguishing between male and female authors. And indeed it is. (Koppel et al., 2002, p. 409).

The machine-learning algorithm they used was able to distinguish between fiction and non-fiction 98% of the time. Interestingly, Koppel et al. (2002) left unexplored in their discussion the question of whether the purpose for which an author wrote influenced lexical and quasi-syntactic choices.

3.3.2 The Argamon et al. 2003 statistical study

Like Koppel et al. (2002), Argamon, Koppel, Fine, and Shimoni (2003) used the texts from the British National Corpus and the features described in Section 3.3. Argamon, Koppel, Fine, and Shimoni (2003) examined these features to identify those that would be most useful for distinguishing texts written by females from those written by males. They then presented statistical comparisons of these features to see which varied significantly depending on author gender. What they found was that females used pronouns of all kinds significantly more frequently than males ($p < 0.01$ after analysis with the Mann-Whitney U-test). Females used most particular types of pronouns more frequently, significantly so with regard to first-person singular pronouns (e.g., *I, me*; $p < 0.01$), second-person pronouns (e.g., *you, your*; $p < 0.01$); and third-person pronouns generally (e.g., *she, her*, $p < 0.01$; *they, them*, $p < 0.05$). Males did, however, use *its* significantly more frequently than females ($p < 0.01$). Males used determiners (e.g., *a, the*; $p < 0.01$), attributive adjectives ($p < 0.01$), and prepositions (e.g., *at, before*; $p < 0.01$) more frequently.

After discussing some details of the varied patterns of use by male and female authors, Argamon and his colleagues offered an interpretive framework, based on work of Biber (1995), who defined a stylistic "dimension" of 67 linguistic variables measuring texts on a continuum from "involved" to "informational" (Argamon, Koppel, Fine, & Shimoni, 2003, p. 332). According to the Argamon and Koppel group, involved writing is associated with female authors and displays "interaction between the speaker/writer and the listener/reader, such as first and second person pronouns"; other "prominent characteristics" of involved writing are "analytic negation [negation with not], contractions, and present-tense verbs" (Argamon, Koppel, Fine, & Shimoni, 2003, p. 332). Informational writing is associated with males and exhibits larger numbers of specifiers and particular types of prepositional phrases. Argamon, Koppel, and their colleagues found that the texts they studied exhibited the involvedness and informational distinction identified by Biber, but they also found other markers useful for categorizing texts that Biber had not accounted for.[8]

3.3.3 Limitations of the Argamon/Koppel 02/03 study

The findings of Koppel et al. (2002), Argamon, Koppel, Fine, and Shimoni (2003) were interesting and inspired other researchers as shown below. These studies were subject to at least three significant limitations, however. First, the they did not explain the research construct of *gender* or how it was assigned to the texts in the corpus, relying instead entirely on the editors/curators of the British National Corpus. Second, they did not explain the research construct of *genre* or how it was assigned to the texts in the corpus, again relying entirely on the editors/curators of the BNC. Finally, because of the nature of the corpus used for the study, it is difficult to be certain of the extent to which a single author was responsible for each text, what I call the "single-author problem." These limitations are common to many of the studies considering the question of gender-differences in language use in written communication. Thus, I'll go into some detail regarding them in the following sections.

In this section, I explained that Argamon, Koppel, and their colleagues examined stylistic lexical and quasi-syntactic features of texts from the BNC. The relative frequencies of the attributes described in this section and used in Koppel et al. (2002) and Argamon, Koppel, Fine, and Shimoni (2003) may function as a measure of the stylistic similarities and differences between texts of two classes (whether the classes are author genders or text genres) without regard to content words that are dictated by the topics of the texts. It is on this basis that I have modeled the features of the empirical study in this dissertation on those of these prior studies. I have briefly explained that the class designations on the text instances in these earlier studies are problematic on the

[8] Koppel et al. (2002) also discussed the dimensional framework of Biber (1995), only much more briefly.

grounds that the researchers did not assign them—they were assigned by third parties in a process not disclosed in the study. I have also noted the "single-author problem."

Studies in technical communication before and since the Argamon/Koppel 02/03 study have looked at gender-differences in communicative practices. The Argamon/Koppel 02/03 study has also been joined by other studies that have attempted classification of texts based on author gender, often taking the issue up in computer-mediated communication, perhaps because text corpora are readily available in that context. Some of these studies explicitly referenced Koppel et al. (2002) without using machine learning methods, while others cited it while extending its machine learning methods. These studies have also sometimes used quite different attributes (features or variables) than the Argamon/Koppel 02/03 study. The following sections take up studies by other researchers citing or extending Koppel et al. (2002) and Argamon, Koppel, Fine, and Shimoni (2003) and addressing the question of gender-difference in written texts. As the limitations in the Argamon/Koppel 02/03 study have been common to the research on this topic, the other studies are discussed in the context of those limitations.

3.4 Gender in studies of gender-difference

According to DeFrancisco et al. (2014, p. 3) gender consists in "the behaviors and appearances society dictates a body of a particular sex should perform," structuring "people's understanding of themselves and each other." Regrettably, even this straightforward definition of gender offers bases for numerous questions and challenges. For example, the behaviors that even feature in gender performances vary from society to society; it is unclear to what extent society "dictates" gender performances; and the extent to which gender structures one's understanding of oneself is not certain. Nevertheless, this simple definition is a reasonable beginning point.

This section explores the ways that other researchers have operationalized *gender* as a research construct. That question is of critical importance when researchers use gender as a variable in their projects. They do this commonly: When researchers report results by saying "the men in this study did X," or "the women in this study tended to Y," the researchers are using gender as a variable, and their work should identify what research construct they are measuring (or attempting to measure) with that variable.

Section 3.4.1 explores numerous studies of gender differences in communication and focuses particular attention on methodological questions about the gender construct. It considers, for example, what researchers say they are studying (if they say at all) when they study *gender*. It also explores their descriptions (if any) of their ascription of gender categories to study participants and the textual artifacts associated with them.

Section 3.4.2 then offers three proposals for researchers to operationalize gender as a research construct. In that section, I propose that researchers should (1) avoid using gender as a variable in their work unless it is necessary to answer their research questions; (2) make explicit their methods for assigning gender categories to participants and artifacts; and (3) respect the difficulties of research participants when asking them to self-identify for gender. Finally, it explains the gender construct I implemented in the empirical study in this dissertation.

3.4.1 Making gender operational in other studies

A great many studies have explored gender differences in human language. This brief survey can address only a few of them, and I have selected them to highlight methodological limitations that I propose to remedy in the empirical study in this dissertation. The studies described here are grouped into three categories: those coming from the field of disciplinary, professional, and technical communication; those citing the Argamon/Koppel 02/03 study, but not using the machine-learning methods of Koppel et al. (2002); and those citing Koppel et al. (2002) and using machine-learning methods.

Numerous studies in the field of disciplinary, professional, and technical communication[9] have addressed differences in the communicative practices of men and women. Common methodological limitations of these studies include the fact that they do not actually define what they mean by "gender" or "sex" and that they do not indicate how they ascribed the gender categories to participants in the study. Some also raise the specter of female participants having to assimilate to masculine, dominance-facilitating communicative practices, despite that fact that the studies do not examine the extent to which the professional discourses are more like the everyday discourse of men or women, or perhaps equally different from both.

[9] See page 2 for a definition of this term.

Mary Schuster (then Mary Lay) has been credited with introducing theoretical discussion of gender in technical communication (Thompson, 2004; Smith & Thompson, 2002). In particular, her observations of student interactions in a technical communication class led her to conclude that to the extent there are differences, "men and women should be free to choose their behavior, rather than being assigned a 'masculine' or 'feminine' role" (Lay, 1989, p. 11). Other studies suggested that students may engage in such strategic androgyny in professional communication. For example, Sterkel (1988) performed careful statistical analyses of stylistic characteristics of students in an undergraduate business writing class. She examined 20 characteristics, including sentence length, numbers of qualifiers, intensifiers, and superlatives, and tag questions. Sterkel found no statistically significant differences between the male and female students' writing, but she concluded that the similarity in style could be the result of the female students assimilating to the male communicative culture, as suggested by R. T. Lakoff (2004). She did not study, however, whether the discourse produced by the participants in her study was more like the communicative practices of men "in the wild." Smeltzer and Werbel (1986) also found no significant differences between the writing of men and women, but again concluded that it could be the result of assimilation to conventional professional language. They had graduate students in English rate the writings of each student based on a seven-point scale identifying whether the author used active or passive voice, had a "positive tone" or a "negative tone," was "easily understood" or "ambiguous," etc. The researchers also evaluated the writings using sentence length, Gunning's Fog Index, "number of clichés, number of paragraphs, and total number of negative words"(Smeltzer & Werbel, 1986, p. 45). In all, the authors measured 16 different writing characteristics. Their analysis indicated no significant relationship between gender and any of the 16 characteristics.

Tebeaux (1990) presented the results of an examination of the quality of student responses to a case-study problem. She concluded that work experience was a more useful predictor of work quality than gender; she further concluded that gender differences were more "concerned with the appropriateness of the response to the writing context than with specific, measurable style characteristics—for example, active/passive voice, number of negative words, sentence length, adjectives/adverbs per sentence" (Tebeaux, 1990, p. 28). Allen (1994) considered whether women writing theoretical articles in technical communication exhibited characteristics commonly attributed to authoritative writing, which she acknowledged as coming from a tradition favoring objectivity. She acknowledged that some of these characteristics—including avoidance of first-person pronouns, hedges, and questions—are elsewhere described as being associated with the masculine communication style.

In addition to Schuster's early study (Lay, 1989), several other studies of communicative differences varying with gender have focused on interpersonal communication rather than on differences in written texts. Rehling (1996) and Raign and Sims (1993) examined language use in interactions of collaborative working groups. Sotirin (2000) considered the role of 'bitching,' a communicative practice often attributed to women office workers consisting of "informal, intimate, opportunistic talk expressing intense emotions like anger and indignation," concluding that "interactional features of bitching enact a micropolitical struggle within and against sociohistorical relations of dominance and oppression" (Sotirin, 2000, p. 24). Other such studies include Brown and Burnett (2006), Tong and Klecun (2004), Wolfe and Alexander (2005), and Wolfe and Powell (2006, 2009).

None of these studies explained what they hoped to measure with the gender or sex construct they used, and none of them explained how gender or sex categories were ascribed to participants. It is likely that the researchers in these studies simply performed informal best-guess assessments of participants' genders based on their common social experiences.[10] One challenge facing all the studies described above that look for textual characteristics that vary with author gender is the challenge of analyzing a large-enough sample of a large-enough number of texts to be able to make generalizations about the results; in the alternative, they face the challenge of statistical power, having a large-enough sample to conclude that any lack of difference is not just the result of sample size.

The view of some researchers in technical communication, including Sterkel (1988), appears to be that communicative agents have two choices: assimilate to arguably masculine, dominance-facilitating communicative practices or be judged as failing to meet professional standards of communication. However, these researchers generally have not performed studies to show whether the discourse produced by the professional (or

[10] See the discussion about this approach in Section 3.4.2 below.

apprentice professional) participants in their studies were more like the communicative practices of men "in the wild." Even if that surmise should be supported by empirical study, CPR theory may offer a means of theorizing and operationalizing resistance to dominance-facilitating communicative practices. CPR theory predicts that successfully varying from genre conventions requires a cost/benefit analysis, an effort to position one's communication as having optimal relevance for the audience. Thus, the Writer's minor departures from conventional expectations should be accompanied by assurances to Readers that at least small marginal advantages will accrue to the Readers if they make the effort to interpret the texts. As long as the Writer keeps the additional effort required to interpret the genre "violation" small enough and offers some additional effect for the Reader (in terms of advancing their goals, for example), the Writer can gradually work to redefine the convention. Over time, such acts of resistance could function to redefine conventions surrounding gender and genre performances. See the discussion of that possibility in Chapter 7 at page 209.

Gender-difference studies are common outside of technical and professional communication. Among the studies citing the Argamon/Koppel 02/03 study but not using its methods is Herring and Paolillo (2006), which set out to assess whether "male blog authors write differently from female blog authors" and whether "[a]uthors of diary blogs write differently from authors of filter blogs" (p. 444). In this study, the researchers concluded that frequency of many word types correlated with genre but few with gender. Herring and Paolillo selected a corpus of blog entries from two blog genres and two genders. In all, they selected 127 entries from 44 blogs, 65 by women, and 62 by men; the average length of each entry was around 281 word tokens. For each text/entry, they assessed the ratio of occurrences of each of several selected lexical features (function-word types) to the total number of words in the text/entry. Though the authors made specific reference to the methods of the Argamon/Koppel 02/03 study, they did not use machine-learning algorithms in their analysis. Nor did they perform part-of-speech tagging on their texts or count bigrams and trigrams as features. Instead, they considered each function-word type's relative frequency separately to determine whether it correlated with author gender or blog genre. Herring and Paolillo (2006, p. 445) assigned gender to blog authors "by examining each blog qualitatively for indications of gender such as first names, nicknames, explicit gender statements. . . and gender-indexical language."

Other studies have used machine learning techniques similar to those used in Koppel et al. (2002). They have commonly used computer-mediated communication, especially blog posts and Twitter feeds, to obtain corpora of data, probably because such data sets are readily available. X. Yan and L. Yan (2006) used a Naive Bayes classifier to classify blog-author genders, achieving an observed agreement of 0.72. They examined blog posts by 3000 authors on a single blogging platform—Zanga.com—concatenating each blogger's posts and treating them as a single instance. The features they used were a *bag of words*, the occurrence frequency of the word types in the corpus, as well as blog background color, fonts, punctuation marks, and *emoticons*. They used blog profile account settings to ascribe gender categories, but they did not describe how frequently account holders indicated their own genders, what gender options were possible, or whether they accounted for account holders posing with genders other than their own.

Argamon, Koppel, Pennebaker, and Schler (2007) showed that MLAs classified blogs successfully based on author gender between 79.3% and 80.5% of the time (observed agreement). They examined blogs on the Blogger.com platform by concatenating the posts in each of 19,300 blogs to create an instance for study, resulting in a corpus of 140 million word tokens. They applied two linear-model machine learning algorithms—Bayesian multinomial logistic regression and multi-class balanced real-valued Winnow—to a feature set consisting of 377 function words and 1,000 common content words to attempt to model blog-author gender and age. They assessed the gender and age of bloggers based on their blog account user profiles. Again, they did not describe how frequently account holders indicated their own genders, what gender options were possible, or whether they accounted for account holders posing with genders other than their own.

Rao, Yarowsky, Shreevats, and Gupta (2010) examined Twitter posts ("tweets") using three different feature sets, the most successful of which obtained 0.72 observed agreement with the gender categories they had ascribed to the texts' authors. Each Twitter user's tweet stream was concatenated to create a single text functioning as an instance for machine learning. Rao and colleagues ran their tests using three attribute or feature sets: One consisted of the lexical unigrams and bigrams in the instances,[11] resulting in more than a million attributes. The other feature set consisted of attributes that the researchers described as *sociolinguistic variables*, mostly consisting of the graphical representations of speech events. For example, they treated individual emoticons as features, as well as instances of "LOL," "ROFL," etc., resulting in 3,774 features. The third feature set—which the researchers called their "stacked model" (p. 41)—consisted of the predictions from the first two models. To each of the abstractions of the corpus of instances into the feature sets, Rao et al. applied a support vector machine (SVM), which generates a hybrid of a linear model and instance-based learning (Witten et al., 2011), to classify the test set. Their sociolinguistic feature set resulted in classifications with an observed agreement of 71.8%; 68.7% for the n-gram feature set; and 72.3% for the stacked model. They identified 1,000 Twitter users and inferred their gender based upon a heuristic: "For gender, the seed set for the crawl came from initial sources including sororities, fraternities, and male and female hygiene products. This produced around 500 users in each class" (Rao et al., 2010, p. 38).

Generally, the studies discussed in this section included very careful descriptions of their methods of data collection and analysis. Though each purported to tell us something about gender, however, they generally did not say what they meant by "gender" or how they ascribed the category. Argamon and his colleagues did not address the complexities of gender assessment. They noted that the works they selected from the British National Corpus (BNC) were labeled for author gender, but they did not indicate how that labeling was done. Herring and Paolillo (2006) assigned author-gender labels to weblog postings "by examining each blog qualitatively for indications of gender such as first names, nicknames, explicit gender statements. . . and gender-indexical language" (p. 445). They offered no theoretical foundation for their qualitative examinations, nor did they provide the means for readers to assess their qualitative heuristics. Smeltzer and Werbel (1986), Sterkel (1988), Tebeaux (1990), and Allen (1994) discussed author genders without indicating how they assessed or assigned them. Argamon, Koppel, Pennebaker, and Schler (2007) and X. Yan and L. Yan (2006) used account settings on blog profiles to assign gender. Rao et al. (2010) used automated heuristics to code

[11] Unlike the n-grams illustrated in Figure 3.1 above, *lexical* unigrams and bigrams are the actual word types, not their parts of speech. So for example, the first three lexical bigrams in the sentence in Figure 3.1 would be MY_AUNT, AUNT_'S, and 'S_PEN.

author gender. None of these studies gave a theoretical account of author gender before applying such labels to texts.

A stand-out exception to the tendency not to explain the ascription of gender categories is Janssen and Murachver (2004), who ascribed author gender using the Bem Sex Roles Inventory (Bem, 1974) and other instruments developed by social psychologists to assess author gender. Despite the challenges with using an instrument like the BSRI, which I discuss in Section 3.4.2, I applaud Janssen and Murachver (2004) for meticulously describing how they ascribed gender categories to participants.

Gender is complicated, and any research classifying texts based upon it should take account of underlying assumptions about it. This study addresses gender by using participant self-identification, but as the next section notes, even this approach leaves some questions unaddressed.

3.4.2 A framework for operationalizing gender

As I suggested in Section 2.2.1, I embrace empirical work that permits at least cautious generalization. In that context, internal and external validity (or validity and reliability) of research findings are ethical concerns (Breuch, Olson, & Frantz, 2002; MacNealy, 1998). I contend that not being explicit about the ascription of the category *gender* as a variable brings into question internal and external validity of research findings. For many of the studies discussed in this dissertation, gender is a variable that is central to the authors' research questions. For many other studies, participant gender is collected as a matter of course, often in the "demographic" portion of an interview or survey. This suggests two different strategies, and based upon them, I will make three proposals here as a framework for the use of a gender variable or construct in studies of written communication.

Three proposals

I offer three proposals: Researchers should (1) avoid using gender as a variable in their work unless it is necessary to answer their research questions; (2) make explicit their methods for assigning gender categories to participants and artifacts; and (3) respect the difficulties of research participants when asking them to self-identify for gender. The first recommendation addresses the choice in strategy I mentioned above: If gender is not a central issue in the researcher's project, she should consider not collecting information about gender. The other two recommendations relate to those studies where the researcher decides to ascribe gender categories.

First, researchers should avoid using gender as a variable unless it is necessary. If a researcher unreflectively collects gender in a survey, the risk is that gender will be used as a "throw-away" cross-tabulation; because it is easy to produce a cross-tabulation of survey responses on a question based on participant gender, it is tempting to share findings using gender as a variable. But in many (if not most) cases, the researcher will not have reflected on what "gender" means in the context of the study. Not collecting gender data poses no significant difficulties in most cases. For example, Larson, Pigozzi, and Lazaraton (In preparation) and Pigozzi, Larson, and Lazaraton (2014) reported the results of a study regarding the perceptions of students of technical and professional communication courses at a large university. They did not collect or report gender information about study participants, because gender was not a variable of importance in their study. They did, however, collect information about students' linguistic and national backgrounds, which were of importance in answering their research questions. This approach is of course easiest where, as with Larson et al. (In preparation) and Pigozzi et al. (2014), findings are reported in the aggregate, without many references to particular respondents.

My second proposal is that researchers who use gender as a variable must make the gender-ascription process explicit. This requires that the researcher adopt a construct definition for gender; that is, the researcher must answer the question, "What does 'gender' measure?" The researcher can either chose a definition of "gender" from an existing theory, or she can identify what she means by "gender" by defining it herself. The researcher should then explain what empirical source permitted the researcher to ascribe the gender category. There are several choices here. Researchers have probably very commonly ascribed gender to study participants based on the researcher's own best-guess assessments: The researcher interacts with a participant and concludes that she is female or he is male. This approach will not likely go away; but the researcher should consider at the time of study design whether and how she will do this.[12] When

[12] I argue that the same goes for other categories. For example, how does the researcher know whether a student in a study is "African American" or "African"? A colleague of mine reported that she had a student whom she described as a "Generation 1.5" American, meaning that though his parents

reporting her findings, the researcher should acknowledge that this is the approach she has taken.

A related approach makes sense where the researcher is studying how participants behave toward each other based on what they perceive each others genders to be. For example, if studying whether a teacher treats students differently based on student genders, the researcher may need to know what genders the *teacher* ascribes to students. The researcher should give thought to how she might collect information about this category ascription from the teacher. The process could prove challenging if the researcher and teacher operate in an environment where students challenge traditional gender roles or where students outwardly identify as transgender.

But participant self-identification should be the gold standard for ascribing gender categories. Except in circumstances where one might not expect complete candor, one can count on a participant to say what her own gender is. This approach to ascribing a gender label respects the autonomy of study participants, as it allows them to assert the gender with which they identify. On the other hand, it does not account for the fact that each study participant may have a different conception of gender, its meaning, its relation to sex, etc. A 76-year-old woman who has lived in the United States her whole life may have a very different conception of what it means to be "female" or "feminine" than does a 20-year-old recent immigrant from Germany. On the other hand, each may be attempting to make sense of her identity as including a female or feminine gender.

In theory, the researcher could address the concerns regarding participant self-identification using a gender role inventory. In fact, one study looking for gender differences in writing did exactly that, using the Bem Sex Role Inventory (BSRI) to assess author genders (Janssen & Murachver, 2004). The challenge with these approaches is that gender is a moving target. Sandra Bem introduced the BSRI in 1974 (Bem, 1974). It has since been criticized on a wide variety of grounds, but of importance here is the fact that it was based on gender role stereotypes from the time when it was created. Blanchard-Fields, Suhrer-Roussel, and Hertzog (1994) described the BSRI this way:

---

were natives of a West African nation, he grew up in the U.S. and spoke English without an accent, at least so far as my colleague could tell. The student firmly identified himself as "African" and *not* "African American," however.

> The BSRI contains 60 descriptive adjectives that individuals rate on a 7-point Likert-scale (1: *Never or almost never true*; 7: *Always to almost always true*). Respondents were instructed: "Below is a list of words that could be used to describe an individual. Please indicate in the space next to each word the degree to which you believe that word describes you." (p. 428).

The adjectives associated with masculinity include "independent," "athletic," and "analytical," and those associated with femininity include "cheerful," "loyal," and "sympathetic" (p. 429-30). But a meta-analysis by Twenge (1997) of studies using the BSRI showed that the masculinity score of women taking the BSRI had increased steadily over 15 years, and men's masculinity scores showed a steady decrease in correlation over the same period. These developments make sense in the context of a gender roles inventory that is necessarily validated over a period of years after it is first developed, resulting in an outdated set of gender stereotypes being embodied in the test. I don't mean to contend here that these inventories have no value for some applications; rather, researchers using them should explain *that* they are using them, *why* they are using them, and what their limitations are.

My third and final recommendation is that the researcher take pains to recognize differences and difficulties that study participants may face in ascribing gender to themselves or to other study participants. For example, assuming that the researcher is collecting participant demographic information with an online survey, she might offer respondents two options for gender: "male" and "female." In contemporary American college classrooms, it's not unusual to have students who do not easily identify with one gender or another or who actively refuse to be classed in a particular gender. Others are confidently transgendered or intersex. Thus, two options may not be enough. However, the addition of an "other" might seem degrading or insulting to one who does not consider herself to be "male" or "female." Another option might be "none of the above," but this again seems to function as an othering selection. There are so many ways that folks might choose to describe their genders that listing them might also be impractical, especially as the list itself might have reactive effects by drawing special attention from the participant to the gender question. Such effects might arise if the comprehensive nature of the list tips participants off that gender is an object of study in the research. I suggest below that a "free-form" space for participants to describe their genders is best for some applications, though it, too, comes with difficulties.

In this section, I've identified and briefly argued for three recommendations when researchers are considering collecting gender information, ascribing gender categories to participants and artifacts, and reporting findings including participant genders. In short, researchers should not collect gender information or ascribe gender categories unless necessary to answer their research questions; if they do ascribe such categories, they should reflect on the best means for doing so in their studies and make the selection explicit when they disseminate findings. Researchers should be explicit about what they mean by "gender" as well. And above all, researchers should show respect for participants and potential participants by avoiding data collection approaches that other them. In the next section, I will describe the approach I have taken in this dissertation.

The gender construct in this study

In this study, I am using methods of text linguistics and natural language processing to assess whether the writing of authors exhibits differences that vary with their genders. For purposes of this study, I define "gender" as as set out in (1).

(1) Gender is a loosely and culturally defined set of social behaviors that are expected to make it possible to distinguish the two most common sexes from each other.

Sex itself can be difficult to define: It's not merely a biological characteristic that can be mechanically assigned, as the features that make up biological sex vary along continua. But we can identify a bimodal distribution: two groups of humans who closely resemble each other within-group and differ between groups—we might call them Sex F and Sex M for convenience. Individuals who do not neatly fit these two sex categories are sometimes described as "intersex"—it is possible that a culture might have no category name for these individuals.

Gender in sexual animals is an important adaptive characteristic, tied with sex, reproduction, and survival. Selecting an individual of the opposite sex and interacting with that individual in certain ways is essential for passing one's own DNA to offspring, who presumably are constrained by the same need. Even in species and cultures, such as the contemporary United States, where it is widely acknowledged that some individuals are predisposed to seek relationships with others of the same sex and not reproduce biologically, the gender categories that organize members of the culture based upon their sex are very important from an early age. Consequently, gender categories are of such pervasive importance to members of human societies that it's difficult to interact without reference to them.

Note that under the gender construct I have defined above in (1), any response by a study participant to a question about his or her gender or sex is regarded as evidence not of his or her sex but of his or her gender; it is a verbal performance, which is to say a "social behavior." Rarely does the researcher have access to the sex of study participants absent a physical inspection, something unlikely in most writing studies projects.

In Section 2.5, I provided an account of a cognitive pragmatic rhetorical (CPR) theory. Section 6.4 provides a relatively detailed CPR-theoretic account of gender to demonstrate CPR theory's utility in explaining the findings in the empirical study in this dissertation. I will summarize it briefly here: I speculate that a cognitive bias in humans makes it a goal of high accessibility and commitment for an agent to ascribe a sex category to other agents as soon as possible in their interactions. From a CPR-theoretic perspective, then, assumptions about a Speaker's sex have a high effect-value to the Hearer. I speculate that each Speaker has accessible a goal to which she is similarly committed: that of signaling her sex by means of gender performance. The gender performances of a Speaker allow Hearers to infer assumptions about the Speaker's sex. Generally, this involves very low effort on the part of the Speaker and Hearer: Because rightly or wrongly (mostly wrongly) certain social roles are allocated to and associated with persons based on their gender performances, and these roles begin to be allocated at early ages, Speakers' gender performances are likely some of the most deeply and habitually ingrained and unconsciously produced in their communicative repertoires. They come with low search costs. Hearers' interpretations of expected performances are also similarly accessible to Hearers. High effects and low effort make gender a relevant category in most cases. These speculations (if true) and facts help to explain the results in this study—where there appeared to be no patterns of gender difference—and to situate them among other studies—where patterns of gender differences were common.

I collected the gender self-identifications of potential participants through a survey that asked a series of demographic questions, one of which was the participant's gender. In order to respect the potential participants and the wide range of possible responses, I chose to offer them the question "Gender?" followed by an open box allowing a free-form response.[13] In this way, I did not constrain the choices that study participants made with regard to identifying their genders. Of 197 study participants, 193 provided information about their genders—the other four not respond to the question. Though this approach respected the possibility that participant genders could vary, it also resulted in a proliferation of different responses. I explain how I addressed them in Section 4.2.3.

In this section, I have explained how the research construct described as *gender* has been deployed in other studies, how I argue it should be deployed, and how I have done so in the empirical study in this dissertation. Gender is the central variable of interest in this study, which considered whether authors' lexical and quasi-syntactic choices varied with their genders in their writing. But the study in this dissertation also set out to hold another variable—genre—constant. The next section describes treatment of that research construct in this study and others.

3.5 Genre in studies of gender-difference

This section explores the use of *genre* as a research construct and variable in empirical research. Of course, the point of the empirical study in this dissertation is to hold the *genre* variable constant to observe the relation of the *gender* variable to writers' lexical and quasi-syntactic choices.

The first part of this section includes subsections that provide an overview of genre theory in disciplinary, professional, and technical communication. Section 3.5.1 offers competing definitions of "genre"; I will offer my own definition in Section 3.5.6. Section 3.5.2 describes the reasons scholars say that genre should be an object of study.

[13] I'm grateful to Dr. Christina Haas for suggesting this solution.

Section 3.5.3 gives a brief overview of methods for exploring genre. And finally, questions and problems associated with genre theory appear in Section 3.5.4. Finally, in Section 3.5.6, I explain how I have operationalized genre in the empirical study in this dissertation.

The second part of this section, beginning with Section 3.5.5, explains the extent to which genre appears to have been taken up in previous studies of gender difference.

3.5.1 Defining "genre"

Genre studies in rhetoric can be traced at least as far back as Jamieson and Campbell in the mid-1970s, but they grounded their thinking in Bitzer in the 1960s and to Burke's conception of "symbolic action." Campbell and Jamieson (1978) defined a genre as "a group of acts unified by a constellation of forms that recurs in each of its members"; "genres are groups of discourses which share substantive, stylistic, and situational characteristics" (p. 20). Campbell and Jamieson saw genre study as an overtly critical effort in their joint work (1978) and in earlier work of Jamieson (1975): They employed the study of genre to assess the effectiveness of generic communication. In their joint work (1978), they discussed Bitzer (1968) at length, and it appears likely that they adopted some of his emphasis on "objective" characterization of rhetorical situation and the possibility for a failure of the rhetor to move her audience. They discussed genre in terms of forms that remain stable over long periods, hundreds or even thousands of years, and they contemplated the possibility that a genre may be repurposed for a new rhetorical situation. For example, Jamieson (1975) explored the example of the early U.S. Congress adapting a form of address to the president from a genre previously used by the English parliament to respond to the king; she judges the choice a failure.

Where Campbell and Jamieson embraced the notion of a genre existing objectively outside of speakers and listeners, Miller (1984) purported to take a view of genre as "social action." In Miller's view, genre is "a conventional category of discourse based in large-scale typification of rhetorical action" (p. 163). She conceived of this level of categorization taking a place in the hierarchy of social meaning-making between "forms of life"—"the cultural patterns. . . that give significance to actions"—and the "episode"—"a rule-conforming sequence of symbolic acts generated by two or more actors who are collectively oriented toward emergent goals" (p. 160).

The early work of Swales (1990) focused on English for Academic Purposes, with particular emphasis on teaching non-native speakers of English, most of them graduate students, to navigate EAP as it is used in the U.S. academy. He emphasized linguistic characteristics, which is consistent with his discipline of applied linguistics. He defined genre as comprising "a class of communicative events, the members of which share some set of communicative purposes" (p. 58). For Swales (1990), genres were the properties of discourse communities, though he did not address the complexities of defining "communities." Later, Swales (2004) reassessed the definition of genre; rather than reaffirming or repudiating his 1990 definition, he supplemented it with a number of metaphors, which he offered almost as lenses for viewing genre.

Berkenkotter and Huckin (1994) provided perhaps the most nuanced definition of genre, which they said is undergirded by theoretical principles of dynamism, situatedness, form and content, duality of structure, and community ownership.

3.5.2 Rationale for studying genre

These scholars offered a variety of reasons for pursuing the study of rhetorical genre. Of those named above, Berkenkotter and Huckin (1994) were perhaps the least explicit about their rationale for the study of genres, the implication perhaps being that such knowledge is interesting on purely theoretical grounds. Each of the others, by contrast, offers a more or less specific rationale. Campbell and Jamieson (1978) and Jamieson (1975) proposed to use genre as a tool for rhetorical criticism, and their research emphasized evaluations of rhetorical performances that used or borrowed generic characteristics. Swales (1990) saw genre knowledge as a tool for educating graduate students, particularly non-native speakers of English, as members of disciplinary discourse communities.

Miller (1984) offered the most sweeping appraisal of the pedagogical implications of genre knowledge: Her view of genre "suggests that what we learn when we learn a genre is not just a pattern of forms or even a method of achieving our own ends. We learn, more importantly, what ends we may have: we learn that we may eulogize, apologize. . . . We learn to understand better the situations in which we find ourselves and the potentials for failure and success in acting together" (p. 165).

3.5.3 Methodological options for exploring genre

These scholars approached study and analysis of genre from different methodological vantages. Jamieson and Campbell's early work focused on textual rhetorical analysis with a consideration of the historical context. For recognizing genres, Campbell and Jamieson (1978) discussed both deductive methods (based on a priori theoretical classifications, etc.) and inductive methods, which do not presume the presence of a genre. So Jamieson (1975) considered contemporary papal encyclicals in light of Roman imperial and early church history. Her description of the genre of contemporary papal encyclicals included references to complex syntax, use of the Latin language, imperial protocols, and "an apostolic salutation and exhortations" (p. 410); Jamieson offered reasons for these choices that are grounded in political realities of the Roman empire. Similarly, she placed the early U.S. presidential speeches and Congressional responses to them in the historical context of 17th and 18th century royal speeches to parliament and parliamentary responses.

The early work of Swales (1990) focused on textual analysis measuring characteristics perhaps best associated with linguistic "register"—lexical and syntactic choices. He did develop an approach for identifying multi-sentence (rhetorical) moves in texts. And to an extent, he included consideration of a social context (which is almost essential to identifying rhetorical moves), but his research methods were lodged firmly in the text.

Berkenkotter and Huckin (1994) by contrast explored the social context for generic texts. Their methods included a case study of a graduate student learning disciplinary genres, interviews with authors of generic texts, and interviews of readers of the same and similar texts, both from within the originating discipline and from neighboring disciplines. Swales (2004) acknowledged this wide variety of methodologies available for exploring genre and urged the use of all of them. The methods should, of course, be driven to some extent by the scholar's purpose in exploring genre.

3.5.4 Questions and problems

A number of problems, questions, and perhaps inconsistencies exist in these views of genre. For example, scholars have considered to what extent genre constrains communication or makes it possible. According to Campbell and Jamieson (1978), "[e]xternal factors, including human needs and exposure to antecedent rhetorical forms, create expectations which constrain rhetorical responses. But the internal dynamic of fused elements also creates expectations which testify to its constraining force" (p. 21; probably echoing Bitzer, 1968). At the same time, other authorities, especially Swales (1990), considered the necessity of convention in all human communication: All linguistic performances rely on conventions for their communicative effect; and in this sense, genres might be thought to create the *possibilities* for communication.

These researchers have considered how stable genres are. Campbell and Jamieson (1978) and especially Jamieson (1975) appeared to contemplate genres capable of stability over decades or centuries. They claimed: "Thus a student who generalized from a sample of 19th century eulogies to the conclusion that eulogies are stylistically florid would be told that a characteristic of the 19th century rhetoric has been mistaken for a generic characteristic and would be urged to sample eulogies from different periods" (Campbell & Jamieson, 1978, p. 22). This quote represents the notion that while rhetorical convention may have changed since the 19th century, there may be a genre of eulogy that has remained stable over the period. It is unclear whether Campbell and Jamieson actually endorsed this view, but they did use metaphors comparing genres to constellations and DNA, suggesting more fixity and constancy (pp. 24-25). Berkenkotter and Huckin (1994) on the other hand overtly embraced "dynamism" as their first theoretical desideratum of a genre: In discussing dynamism, they referred to the characterization by Miller (1984) of recurrence "as an intersubjective phenomenon, a social occurrence"; but Berkenkotter and Huckin argued that "recurring situations resemble each other only in certain ways" and that genres are thus "always sites of contention between stability and change" (6).

One final interesting point of discussion is whether and how genres relate to discourse communities and to authorial agency. Campbell and Jamieson (1978) and Jamieson (1975) exposed genre use to criticism based principally on its communicative or rhetorical effectiveness, consistently with Bitzer's assertion that the rhetor is operating in an objective situation with an exigence and constraints, but with some individual goal for using rhetoric to change the situation. Swales (1990) early embraced authorial agency: he claimd genres have a "double generative capacity"—"to establish rhetorical goals and further their accomplishment.” Later (2004), he acknowledged some of the complication of questions of authorship and agency in contemporary philosophy of language. Miller’s pedagogical rationale for genre study (see above) expressly identified “potentials for failure and success in acting together” (165), which seems to carry the concept of agency into the disciplinary or discourse community. Berkenkotter and Huckin warn, though, that “asserting a relationship between the concept of genre and that of ‘discourse community’ is a slippery proposition because neither concept refers to a static entity” (p. 21). Nevertheless, they referred somewhat approvingly to Swales’ (1990) conception of discourse community.

Cognitive pragmatic rhetorical (CPR) theory may well help to explain the functioning of important theoretical processes in genre theory, including the stability and dynamism of genres and the process by which certain situations become typified. Stability responds to audience assumptions and goals, and thus to minimizing processing effort; the Writer can count on the audience adopting the easy interpretation of a work that conforms closely to genre conventions because the audience’s goals, assumptions, and habitual interpretive practices will guide it to that interpretation. But assume for the moment that the Writer cannot achieve with the current genre conventions what she wants to achieve. Consider a company employee who prepares a periodic report that has genre status within the company. Perhaps another communicative approach would cost her less effort, better highlight her own contributions, allow her to take credit for saving the company money, gratify her desire for efficiency, or offer her some other benefit. Any effort she makes to vary from the existing genre conventions is likely to be opposed with the cumulative force of other individuals to the extent that they individually perceive the increased cognitive effort of understanding the new report exceeds the effects associated with it. Our hero can predict this resistance (thanks to the operations described by CPR theory) and revise her strategy to address the cognitive effort/effect balance.

Genre theory has so far done too little to understand the workings of “typification.” Miller (1984) purported to adopt the typification theory of German sociologist Alfred Schutz.[14] This theory emphasizes the potency of language and social practices in typification. We can illustrate this with an anecdote from Bawarshi and Reiff (2010):

[14] There is a more extended discussion of Schutz’s typification in Section 2.4.2.

At a music festival, Bawarshi's daughter saw a little boy in a princess costume; she refused to accept that the boy was not a girl, because of the gender convention (princess costume) that he had adopted. Children clearly see gender conventions enacted every day, and so their cognitive environments include habitual attributions of gender to other persons based on those gender conventions. But this does not illustrate how knowledge of communicative types arises in individuals and can do so based even on single events. Consider a contrasting anecdote about a car purchase from Schank and Abelson (1977), recounted by Roger Schank:

> My daughter Hana (age 4) was with me when we bought it and asked if I was going to get a new key chain. I asked her what she meant. She replied that when we had gotten our old car in Rhode Island (where it had arrived off the boat 2 years earlier) I had bought a new key chain. This was her only experience with getting a car and already the events in it were a script for her. (p. 68)

What cognitive processes resulted in Hana typifying the one experience she had previously had of buying a new car? What processes resulted in her picking out the corresponding acquisition of a new key chain as part of the script for getting a new car? There are, of course, analogous questions in genre theory, genre theory should seek the answers to these questions, and CPR theory offers a framework to help find them.

In the empirical study in this dissertation, I have ascribed a genre variable to the texts provided by study participants, and as I claim below, that variable is held constant in this study: That is, the participants all wrote texts in the same genre. This fact will be important for explaining the findings of the study, taken up later. But before explaining the genre construct I have employed in this study, I would like to describe the extent to which the previous studies I introduced in Section 3.3 and Section 3.4.1 made use of genre constructs.

3.5.5 Making genre operational in other studies

This section considers the extent to which studies of gender differences in communication discussed above implemented any theoretical genre construct or attempted to control for genre. It first reviews Argamon, Koppel, Fine, and Shimoni (2003) and

Koppel et al. (2002), the studies I have referred to collectively as the Argamon/Koppel 02/03 study, which are described in great detail in Section 3.3. It then considers studies in disciplinary, professional, and technical communication, which were described in Section 3.4.1. Finally, it reviews studies citing the Argamon/Koppel 02/03 study, also originally described in Section 3.4.1.

The Argamon/Koppel 02/03 study made reference to the genres of the texts the researchers studied. But in that study, the researchers relied upon the ascription of genre categories by the British National Corpus—the principal genre categories being "fiction" and "non-fiction." Though these *are* kinds or types of texts, thus warranting the use of the term "genre" in the broadest sense, they do not constitute genres in the genre-theoretic sense. The BNC includes texts published as newspaper and magazine essays and books, without regard to the authors' intended audiences, without regard to the authors' intentions for publishing their works, etc. Absent a narrower context, it would be difficult to describe these texts as consisting of a single genre. There are undoubtedly many genres represented in the BNC, but "fiction" and "non-fiction" are not detailed enough to satisfy any of the genre conceptions above.

Of the studies described in Section 3.4.1, those arising in disciplinary, professional, and technical communication come closest to being controlled for genre. Some did not address gender differences in the texts produced by participants, so I will not consider them here (Brown & Burnett, 2006; Lay, 1989; Raign & Sims, 1993; Rehling, 1996; Sotirin, 2000; Tong & Klecun, 2004; Wolfe & Alexander, 2005; Wolfe & Powell, 2006, 2009). But others resulted from the efforts of students writing assignments in a class (Smeltzer & Werbel, 1986; Sterkel, 1988; Tebeaux, 1990). As I explain in Section 3.5.6, I consider student writing to have genre status when students write in a classroom context where all are being directed to create a text of a conventional form. Perhaps not surprisingly, these studies are the ones that appeared least likely to find gender differences. In other words, learners of a new disciplinary or professional discourse intent on producing texts conforming to disciplinary or professional genre conventions were unlikely to write in ways that varied based on their genders. One other study in this field looked only at essays written by women appearing in scholarly journals of one kind; in effect, it held both gender and genre constant (Allen, 1994). But as that study did not explore gender differences in texts, it is of no particular import here.

A great many of the other studies of gender difference above gathered texts that could not in any way be described as being of a single genre. For example, Herring and Paolillo (2006) defined key genre and sub-genre terms they used in their article: A "weblog" or "blog" as "publicly-available websites, typically single authored, in which dated entries are posted in reverse chronological sequence" (p. 440). The authors coded sub-genre or "blog genre" as "diary," which provides "report and comment on the author's own life," or "filter," which describes "events external to the author" (p. 445). But blogs, even diary blogs and filter blogs, do not represent a typified response to a recurring situation. Indeed, though Miller and Shepherd (2004) characterized blogs as a genre, they later recanted, acknowledging that blogs did not fit into the theoretical conception of genre (Miller & Shepherd, 2009).

Other studies of blogs also failed to control for genre, including Argamon, Koppel, Pennebaker, and Schler (2007) and X. Yan and L. Yan (2006). If blogs do not constitute a genre, Twitter posts *a fortiori* do not either. After all, the social purposes for which people tweet are widely varied. Twitter posts are a genre only in virtue of the medium by which they are distributed; in the same way that all spoken utterances taken together constitute a genre. Thus, Rao et al. (2010) and Burger, Henderson, Kim, and Zarrella (2011) cannot be said to have explored texts in a single genre.

In summary, the only studies that appear to have controlled for text genres are studies of the writing of students writing in disciplinary, professional, or technical communication classrooms. In the past, the work of such students has been described as not having genre status or as being "pseudotransactional" (Spinuzzi, 1996). As I shall explain in the next section, the CPR-theoretic conception of genre includes student writing on certain kinds of assignments in some classrooms.

3.5.6 A framework for operationalizing genre knowledge and genres

When agents gather genre knowledge about the social environments in which they interact, they gain them one interaction at a time. The first introduction one has to a new text that is arguably an instance of a genre presents a challenge: Should one interpret it as being a genre and attempt to appropriate its formal characteristics for future communications directed at similar purposes; or should one conclude that this text is a "one off" approach to the goal it addresses and consider it "ungenred"? As one sees further instances of a possible genre, the agent is confronted each time with slight variations, but she must decide which of the variations are genre violations and which are permitted in the context of the genre. She does not do this based only on deductive rules but rather based on repeated interactions with agents around her. In terms common to Miller (1984) and Straßheim (2010), she must decide whether to "typify" the text and the social situation in which it operates and determine which formal characteristics of the text are essential to its social function. Cognitive pragmatic rhetorical (CPR) theory predicts that these interactions will be governed by the principles of relevance, and that the agent will use the CPR-theoretic production and comprehension procedures.

From a Speaker or Writer's perspective, then, genre can be defined as in (2).

(2) A genre is a loosely and culturally defined set of communicative behaviors, usually formal conventions, a Speaker or Writer expects to have a particular effect or effects on a Hearer or Reader, based on assumptions about a typified situation in the Speaker's imputed cognitive environment.

I will discuss this formulation in greater detail in Section 6.5, but I offer some brief conclusions here. First, formal conventions may include linguistic register, which I define after Matthews (2007), as a "set of features of speech or writing characteristic of a particular type of linguistic activity or a particular group when engaging in it."[15] Second, under CPR theory an agent's genre knowledge is just a "slice" of her cognitive environment.[16] An agent's cognitive environment, and thus her genre knowledge, grows out of her understanding—her assumptions—about a recurring, typified situation and the expectations of other persons around her for the appropriate or fitting response to that situation. The agent herself need never have experienced the typified situation to attempt a genre performance. Of course, her text may not meet the expectations of a particular Reader for the genre. This is especially true where the Writer is an apprentice member of a disciplinary or professional community.

This approach to genre knowledge avoids the dangers of reifying genres. Most genres are not fixed things subject to categorical descriptions, just as the meaning of most words

[15] I have avoided the broader definition for *register* from Biber (1995), who appeared to include many other formal characteristics that I would describe as being within the scope of *genre* rather than *register*.

[16] Section 2.5 provides a more complete overview of CPR theory.

is not fixed and subject to categorical description. Though there might be a central tendency among the participants in a communicative exchange about appropriate formal conventions for that exchange, each agent likely has a unique, idiosyncratic imputed cognitive environment. Because genres give typical results in typified situations, agents can and do depart from genre conventions all the time in order to achieve effects. Writers must do so, however, recognizing that if they make the effort for their Readers to interpret the change greater without promising some kind of greater effect, the Writers may fail to achieve the desired change in the Readers' cognitive environment. This is a recognition of the relevance ratio: Relevance increases with expected and desired cognitive effects but decreases with cognitive effort.

In this section, I have described contemporary conceptions of genre in genre theory, noted that those conceptions have generally not been employed in studies of gender differences in communication (even while some studies managed to control for genre inadvertently), and provided my own conception of genre. In the next section, I take up the final methodological limitation of previous studies, what I call the "single-author problem."

3.6 The single-author problem

When considering whether a text exhibits markers of some characteristic of its author, it is important to know who the text's author is. And when the characteristic of interest is one exhibited by individuals, such as gender, it is helpful to know that the text represents the work of that individual, and only that individual. This is perhaps more difficult than it might seem at first blush.

For example, Pakhomov, Chacon, Wicklund, and Gundel (2011) attempted to extend work by Garrard, Maloney, Hodges, and Patterson (2005) examining syntactic complexity and lexical features in Iris Murdoch's writings. The Irish author suffered from Alzheimer's disease before she died in 1999; she was diagnosed after publishing her last novel, *Jackson's Dilemma*, in 1995. Pakhomov, Chacon, et al. (2011) desired to extend the work of previous studies of Alzheimer's sufferers that had examined "various aspects of language production and comprehension including sentence structure complexity, idea density, use of referring expressions and discourse coherence" (Pakhomov, Chacon, et al., 2011, p. 136). Both the Pakhomov and Garrard studies prefaced their expositions with claims about one important issue: How can we know that Iris Murdoch's novels represent the work of Murdoch and not her editors? The researchers acknowledged that publishers' editors often have substantial effect on the final published work. The explanation on which they relied is that Murdoch was legendary for resisting editorial interference in her work. They acknowledged that their studies could not work with authors who might be more subject to the extensive editing of the publisher.

In the studies of texts described in the previous sections, however, only Janssen and Murachver (2004), Smeltzer and Werbel (1986), Sterkel (1988), and Tebeaux (1990) took steps to ensure that the texts being analyzed came from single authors whose genders were identified. The others failed to address this concern (Argamon, Koppel, Pennebaker, & Schler, 2007; Burger et al., 2011; Herring & Paolillo, 2006; Rao et al., 2010; X. Yan & L. Yan, 2006). The study in this dissertation addresses the issue by using texts from first-year legal writing courses at two law schools. These programs required students to do all their writing alone, generally in isolation from each other; e.g., they did not engage in substantial peer review work during drafting. I therefore argue that we can accept these papers as single-authored with confidence.

3.7 Conclusion: Do men and women write differently?

This chapter has provided essential background to explain the motivation for performing the gender-difference study described in this dissertation. Several studies using large (or larger) data sets generally showed there were differences in the writing of men and women (Argamon, Koppel, Fine, & Shimoni, 2003; Argamon, Koppel, Pennebaker, & Schler, 2007; Koppel et al., 2002; Rao et al., 2010; X. Yan & L. Yan, 2006). This outcome is consistent with folk beliefs and popular works (including, for example, Tannen (2001)) that there are deep differences in the communicative styles or cultures of men and women (DeFrancisco et al., 2014). As a consequence, I will assume that in unmarked situations, men and women probably do make different stylistic choices.

At the same time, a small number of earlier studies in professional and technical communication showed no significant differences between the writing of female and male authors (Smeltzer & Werbel, 1986; Sterkel, 1988; Tebeaux, 1990). Each of these studies examined student writing in professional or technical writing courses, and the authors concluded that the lack of gender differences there could be attributable to the students' efforts to "assimilate" to conventional professional language. Unfortunately, these studies had sample sizes that were too small—that is, they had too little statistical power—to infer from the lack of difference in them that there would be a similar lack of difference in a bigger sample or in the population at large. Nevertheless, they suggested that the different communicative styles or cultures of men and women may not be even skin deep, that men and women, when faced with a common task after receiving common training, produce communicative performances that are stylistically indistinguishable.

This dissertation presents a well designed study of a larger sample of student writing prepared in a professional training environment; and it shows that gender differences (if they existed before these students came to law school) did not persist in such writing contexts.

In Section 3.2, I considered the continuing debate about whether researchers should perform research that looks for gender differences in human communication. I considered the arguments of DeFrancisco et al. (2014), who argued that research on gender differences reinforces essentialist views of gender, tends to conflate sex and gender as variables, overlooks substantial similarities in performances between women and men, and does not provide the rich intersectional analysis that ethnographic and other research methods provide. I refuted these arguments with my own and then described three benefits that Hultgren (2008) saw in the kind of empirical research that I describe in Chapter 4. Ultimately, the conflicting views of Hultgren (2008) and DeFrancisco et al. (2014) represent conflicting epistemic commitments, which I believe can be reconciled if researchers value the complementary contributions of quantitative and qualitative research.

Because the empirical study in this dissertation was motivated by and uses the same stylistic features as the Argamon/Koppel 02/03 study (Argamon, Koppel, Fine, & Shimoni, 2003; Koppel et al., 2002), I described it in considerable detail in Section 3.3. The Argamon/Koppel 02/03 study analyzed stylistic features—lexical and quasi-syntactic decisions of writers—to study texts from the British National Corpus. They found variations in frequency depending on author gender (Argamon, Koppel, Fine, & Shimoni, 2003), and they found that machine-learning algorithms could distinguish texts written by women from those written by men around 80% of the time based on those features (Koppel et al., 2002).

Ultimately, however, the Argamon/Koppel 02/03 study suffered from three important methodological limitations, which I described in Section 3.3.3. First, it is unclear how the gender of text authors was ascribed to the texts that Argamon and his colleagues studied. Second, the Argamon/Koppel 02/03 study did not control the texts for genre in a genre-theoretic sense; that is, they did not study texts where the authors shared key components of their cognitive environments, including goals, assumptions about typified situations, etc. And finally, they did not address what I have called the "single-author problem," uncertainty about whether texts are written or entirely written by the person under whose byline they appear.

These three limitations provided the structure for a review of other studies and for me to articulate my own framework for addressing them. In Section 3.4.1, I described previous studies of gender differences in language, including studies from the disciplinary, professional, and technical communication field and later studies citing the Argamon/Koppel 02/03 study or using its methods. I concluded that most of them had failed either to explain their conception of *gender* as a research construct, to explain how they had ascribed gender categories to texts, or both. In Section 3.4.2, I offered my own framework for operationalizing gender as a research construct, including three proposals for researchers considering using gender as a variable. I then defined the gender construct for this study and explained how I ascribed gender categories to the papers in the sample analyzed.

In Section 3.5, I laid out the definitions from genre theory in disciplinary, professional, and technical communication and described the rationale and methods for studying genres in this sense. After describing some questions or problems with genre theory, I noted in Section 3.5.5 that only the disciplinary, professional, and technical communication studies above made any effort to control the samples they analyzed for the writers' rhetorical purpose. I then briefly explained in Section 3.5.6 the genre conception, grounded in CPR theory, that I have employed in the study in this dissertation.

Finally, I described the "single-author problem" more completely in Section 3.6 and claimed that this study resolves it to a reasonable degree of probability.

Two groups of studies—big-data studies that showed gender differences in texts without a common purpose or genre, and smaller studies that showed no gender differences in texts written with a common purpose—prompted me to ask whether gender and genre interact, and if so, how. I wanted to see if gender differences of the kind previously identified would appear in a larger sample of texts that were controlled for text genre and attempt to offer a theoretical explanation for those findings, whatever they might be. That led to the empirical research study in this dissertation, the methods for which are described in Chapter 4.

I did not find the patterns of difference seen in the previous big-data studies, as Chapter 5 shows. In Chapter 6, I demonstrate how CPR theory could explain these findings.

Chapter 4

Study design: Seeking gender differences in genred writing

4.1 Introduction

In Chapter 1, I explained the motivating question underlying this dissertation:

> *How can cognitive pragmatic rhetorical (CPR) theory contribute to our understanding of rhetorical and disciplinary, professional, and technical communication theory and in particular to our theories of gender and genre performances?*

I proposed in Section 2.5 that CPR theory can explain the tension between a Writer's desire to conform to text genres—especially in a new disciplinary environment—and her tendency to engage in habitual linguistic practices acquired when writing in genres learned earlier or in ungenred texts. CPR theory takes account of a Writer's goals and beliefs about the world and explains the efforts the Writer employs in finding, discovering, or inventing her communicative performances with the principle of relevance, which holds that she will expend effort in her writing choices that is commensurate with the accessibility and strength of the goals for which she writes and the effect she expects her writing to have on the cognition of the Readers(s).

In Chapter 3, I discussed previous studies of gender difference in written language. I explained how *gender* and *genre* can be operationalized as research constructs, but I noted that most previous studies suffer from significant limitations arising from their methods for assigning these category variables to texts, and from the "single-author problem." Nevertheless, I concluded in Chapter 3 that the evidence available so far suggests that writers habitually make choices about their writing that vary with their genders, at least in ungenred texts; that is, men and women appear to write differently when they are writing without the constraints of disciplinary genres. There is counter-evidence, and the methodological limitations of some of the previous studies make one reluctant to rely on them. But I also noted in that chapter that folk beliefs support the view that there are deep-seated differences between the sexes in communication. The study here thus makes a presumption that such differences existed in study participants here before they came to law school. See Section 7.2 for a discussion of the ways that future research could be used to interrogate this presumption.

But it is an open question whether gender differences, if they exist in everyday ungenred communications, will endure when writers are trained in the use of disciplinary or professional genres, and if they endure, whether they will take the same form in genred and ungenred texts. The answers are important because they help to address the concerns raised in Chapter 3 that previous studies of gender difference in writing may have essentialized differences between female and male authors. They are also important because they serve to inform understandings of communication and rhetorical theory, particularly the ways that writers may work to leave behind writing habits from their pasts—where they wrote ungenred texts or texts in other genres—to adopt the genre conventions of a new disciplinary or professional community.

This chapter describes the methods of an observational study to answer these questions. Chapter 5 presents the results of statistical analysis and numerous machine-learning trials performed based on this design. And Chapter 6 demonstrates the utility of CPR theory in explaining the findings of this study—that is, whether female and male authors—or more particularly, Gender F and Gender M authors—write differently when engaged in professional communication, and if so, why.[1] Along the way, this study seeks to address some of the methodological limitations of the previous studies.

[1] See the discussion in Section 4.2.3 regarding the choice to refer to the authors in this study as "Gender M" and "Gender F," rather than as "male" and "female" or "masculine" and "feminine."

This is a descriptive, observational study applying the tools of statistics and natural language processing (NLP)—and particularly supervised machine learning—to examine texts produced in several sections of a law school legal writing class at two law schools and to assess whether the language of the texts the students produced exhibited differences that varied with their self-reported genders. The students in these classes prepared a year-end memorandum of law—also called a *brief*—with all the students writing a document in the same genre, and in many cases, on the same hypothetical legal case.

This study is one way of approaching the motivating question above. Executing it requires a decision of how to assess whether texts written by Gender F authors are different from those written by Gender M authors—i.e., what features of the texts will be *counted* when addressing this question. It is also desirable to determine *how to count* those features with a large enough group of texts to make some generalizations, or at least to suggest generalizations. For that purpose, this chapter and the empirical study in it take as a model Koppel et al. (2002) and Argamon, Koppel, Fine, and Shimoni (2003).[2]

Argamon, Koppel, Fine, and Shimoni (2003), which was described more fully in Section 3.3, applied statistics to lexical and quasi-syntactic features of texts to assess gender differences in them. The features used in that study were stylistic characteristics: because they focused on function words and parts of speech, they were generally not measuring the content-orientation of the texts. Statistics have also been used widely in other studies of writing. The value of statistics is in assessing, among other things, whether differences between sets of observations are statistically significant; that is, what the probability is that the observed differences are the result of a real underlying difference between the observed phenomena and not merely the result of chance. One *challenge* of statistical assessments of significance is that they require the researcher (either as an individual or as a member of a disciplinary community of researchers) to decide what the boundary or *threshold* for statistical significance will be for the study.

[2] Throughout this dissertation, Koppel et al. (2002) and Argamon, Koppel, Fine, and Shimoni (2003) and the underlying data set are occasionally referred to as as the *Argamon/Koppel 02/03 study*. Note that Argamon, Koppel, Pennebaker, and Schler (2007) is also cited in this dissertation; this 2007 study, cited and discussed elsewhere in this dissertation, is a different study than the 2002/2003 one and not what is meant in references to the *Argamon/Koppel 02/03 study*.

Koppel et al. (2002) and other studies that have come after it have also used methods from computer science and artificial intelligence called *machine learning algorithms* (MLAs). As we saw in Chapter 3, MLAs are powerful tools for identifying patterns involving many features that can serve as a basis for distinguishing or categorizing texts. The features on which MLAs rely to classify texts are not necessarily different from each other in a statistically significant way. MLAs may thus make it possible to assess differences that are more subtle than those recognized by tests of statistical significance, and they do not rely on the researcher's possibly arbitrary threshold of statistical significance to determine which features are useful for classification. On the other hand, the models produced by MLAs may not be interpretable in the same way as statistical models, if they are interpretable at all. MLAs are also interesting because they have been applied to technical and professional communications to achieve organizational objectives, particularly in medical and legal informatics. See particularly the discussion in Section 4.4.2 and references there to the work of Humpherys, Moffitt, Burns, Burgoon, and Felix (2011), McCart, Berndt, Jarman, Finch, and Luther (2013), Pakhomov, Hanson, Bjornsen, and Smith (2008), and Pakhomov, Shah, Hanson, Balasubramaniam, and Smith (2010). But they have not received attention in the disciplinary, professional, and technical communication literature, an oversight this study begins to redress.

Thus, this empirical study contributes to answering the following specific research questions:

1. Do Gender F and Gender M writers in a disciplinary genre in which they are being trained use lexical and quasi-syntactic stylistic features with relative frequencies that vary in relation to their genders?

2. If so, do the differences appear in interpretable patterns?

3. Can machine-learning algorithms categorize the same texts by author gender based on the same features?

4. If so, do they provide interpretable models?

These answers in turn allow us to explore the over-arching research question in this dissertation, assessing the utility of cognitive pragmatic rhetorical (CPR) theory and how it contributes to our understanding of rhetorical and technical and professional communication theory and in particular to our theories of gender and genre performances.

This chapter explains the methods of this study. Section 4.2 describes the instructional context and data collection for it. In Section 4.2.1, I argue that the classroom context and the writing assignments to which students responded resulted in all the writing samples used in this study being of the same *genre*. (See also Section 3.5 regarding genre as a research construct.) Section 4.2.2 explains that law students in these classes are under intense pressure to produce their work without any collaboration with other writers. Consequently, these writing samples do not give rise to concerns about texts that are written by multiple authors, subject to heavy editorial control, or "ghost written." They are thus not subject to the "single author problem" described in greater detail in Section 3.6. In Section 4.2.3, I contend that this study has sought to pay careful attention to *gender* as a research and social construct, addressing limitations of many previous studies seeking gender differences in language use. (See Section 3.4 for a more extensive discussion.) Section 4.3.1 describes the procedure for collecting data and the resulting data sets.

Section 4.4 reports the process of data preparation and analysis. This includes preparing the texts for the application of statistics and machine learning, covered in Section 4.3.2, and arguments for statistical analysis, described in Section 4.4.1, and the application of machine learning, described in Section 4.4.2. Section 4.5 takes up some ethical issues in this study design.

4.2 Law school context

In the American legal system, lawyers are trained in post-baccalaureate professional schools, usually for three years of full-time study. The pressures that students in these environments feel to conform to disciplinary conventions in general have been explored in popular fiction and memoir, including the novel and television series *The Paper Chase* (Osborn Jr., 2004) and Scott Turow's *One L* (Turow, 2010). Insiders in the legal education industry have sometimes criticized the legal academy for the stress and confusion it imposes on its students (Caulley & Dowdy, 1986). Educators and law students alike acknowledge the aptness of the old adage about law school education:

"first year they scare you to death, second year they work you to death, and third year they bore you to death" (Kahlenberg, 1999, p. 159).

Some researchers have explored students' efforts in law school to function within and conform to the language of the law, both as it is spoken (Mertz, 2007) and written (Cauthen, 2010). These studies have emphasized the challenges that students face and the power dynamics enacted using language in the law school—which are usually presumed to be only a foretaste of the power dynamics of legal language in the courtroom and boardroom.

Based on my anecdotal experiences as a teacher in the legal-writing classroom for eight years, I claim that law students have an intense desire to conform to the disciplinary conventions of the profession into which they are training. These students are mostly very eager and often very bright. They passionately desire to succeed in law school in hopes that it will open doors for the kinds of jobs they imagine they want. In this context, when students are asked to write within recognized genres in their chosen profession, we can expect that they will direct all the effort they can to adhering to the genres' conventions, including linguistic register. Law school calls upon students, regardless of their gender, to leave behind old habits of thought and language and to embrace new ones; the students recognize that their responsiveness to this call may determine their future opportunities.

The data for this study were collected at two law schools in the U.S. Midwest during the 2011-12 academic year. One of these schools, referred to here with the pseudonym "Academy School of Law," is routinely ranked among the top 35 law schools by popular national assessments such as *U.S. News and World Report* and *Above the Law*. The other, referred to here as "Lyceum Law College," is not routinely ranked among the top 100 schools accredited by the American Bar Association. According to the administrations at these two schools, they enrolled a total of 545 new students in AY2011-12; of them, 263 were female and 282 male according to law school records. Each school required as part of its first-year curriculum several basic courses, including contracts and civil procedure. Importantly, each also required students to take a course or combination of courses in legal research, analysis, and writing.

It is in this context that I collected writing samples from 193 gendered authors and created the text corpus that was the object of analysis for the study. The research questions posed above call for texts written by single authors of different genders working in a context where the authors would be attending closely to, and attempting to adhere to, conventions of a single disciplinary genre. As Section 4.2.1 explains, the first-year legal writing classrooms that were the location of this study provide the appropriate context for collecting texts of a single *genre*. Section 4.2.2 shows how the law school context allows this study to escape the *single author problem*, because law students are generally not permitted to work with each other on their school writing. Section 4.2.3 explains why collecting the authors' genders for these texts in a *free-form* fashion results in a better assessment of the author-gender category than the studies discussed in Chapter 3.

4.2.1 Texts in a professional genre

Section 3.5 proposed that *genre* as a research construct is the application of a category label to a set of texts exhibiting a loosely and culturally defined set of communicative behaviors, usually formal conventions, a Speaker or Writer expects to have a particular effect or effects on a Hearer or Reader, based on assumptions about a typified situation in the Speaker's imputed cognitive environment. In the present study, there is evidence that the participants, all students finishing their first year of training in law school, shared certain elements of their cognitive environments, including accessible, though possibly only weakly held, assumptions about the formal conventions of legal writing and of the hypothetical memoranda they were writing—the typified situation; intense and accessible goals to do well in this important assignment; and assumptions about the cognitive environments they imputed to their instructors. I argue that these elements, taken together, show the students in my study were all writing in the same *genre*.

Section 3.5.5 observed that previous studies of gender difference in language have often failed to consider whether language differences exist when men and women write in the same genres. So, for example, studies have examined a broad corpus of published texts including journalistic, fiction, and other published texts (Argamon, Koppel, Fine, & Shimoni, 2003; Koppel et al., 2002). They have considered blog posts generally (Argamon, Koppel, Pennebaker, & Schler, 2007; X. Yan & L. Yan, 2006), or subcategorized as "diary" and "filter" blogs (Herring & Paolillo, 2006). And they have examined Twitter feeds (Rao et al., 2010; Burger et al., 2011). In each of these cases, however, it is difficult to see many shared assumptions guiding the authors, any formal conventions dictated by typified situations, or expectations shared by all (or most) of the authors' Readers. In short, these texts are not of the same *genre*.

The application of the genre construct matters to studies of gender difference in language, because it helps to answer questions about the arguably essential nature of gender differences and about the durability of them, if they exist in the first place. Previous studies, including R. T. Lakoff (2004), Smeltzer and Werbel (1986), and Sterkel (1988), and comments of reviewers of this manuscript have suggested that the women (in previous studies) or Gender F authors (in this study) are expected or asked to abandon their gendered habitual practices when entering male-dominated professions. This raises difficult questions about social power, suggesting perhaps that Gender F authors had to devote more cognitive effort to writing in the law-school genres than did their Gender M peers. I will take this issue up as a topic for future study in Section 7.4, but I would like to start here by noting that I know of no systematic study showing that women must adapt their writing styles more than men in these contexts. My own anecdotal experiences as a teacher of this kind of writing for eight years is that the (apparently) male students (apparently) have as much difficulty as the (apparently) female students. These observations are not systematic, though, and they prompt me to suggest further empirical study.

Use of this genre construct here also illuminates the ways in which acculturation to a professional writing community works for gendered persons; it allows consideration of whether writers abandon gendered linguistic practices when attempting to conform to genre conventions, including linguistic register, and CPR theory would help explain the degree to which this effort is successful. This study addresses these issues by collecting texts that are of a single professional genre. And this section describes how the first-year students at Academy School of Law and Lyceum Law College prepared such a set of texts, first describing the legal writing programs and then the year-end brief or memorandum assignment.

According to officials at at these law schools, the first-year legal writing classrooms at Academy School of Law and Lyceum Law College shared some characteristics and differed in others. At Academy School of Law, students were grouped in 25 sections, with each section having between eight and ten students and each having an adjunct attorney instructor, usually a practicing attorney from the community, and a student instructor, a second- or third-year student acting as an "upper-level student teaching assistant." The syllabus and assignments for the year were controlled from a central legal writing administration. Thus, for the spring assignment that is the object of this study, all the students at Academy School of Law wrote about the same hypothetical problem. Required texts at Academy School of law were Clary and Lysaght (2010) and *The Bluebook: A Uniform System of Citation* (2011). I should note here that I have taught the course that is the locus of this study at Academy School of Law for eight years, though I did not teach it the year that I conducted this study. I have, from time to time, made observations in this dissertation grounded in my intuitions or anecdotal experiences; where I have done so, I have tried to acknowledge the source of those observations and distinguish them from observations gathered by more systematic means.

Lyceum Law College also grouped students into small sections, in its case, 28 sections of nine to twelve students. There, however, each section was taught by a single adjunct professor, again usually a practicing attorney, but with no student teaching assistant. Furthermore, legal writing professors at Lyceum Law College were responsible for developing their own hypothetical problems for students to write about, within certain constraints established by the school's legal writing program. Required texts at Lyceum Law College included Schmedemann and Kunz (2007) and *The Bluebook: A Uniform System of Citation* (2011).

According to administrations at these law schools, each school required students to write a spring capstone assignment, typically an example of what lawyers call a "motion practice brief": The students wrote memoranda of law in support of or opposition to a hypothetical motion seeking dismissal of a claim or summary judgment on a claim. At Academy School of Law, each student wrote a memorandum supporting or opposing a motion to dismiss a hypothetical copyright claim. At Lyceum Law College, students' memoranda supported mostly motions for summary judgment and a few for dismissal; the legal subject matter of these hypothetical cases varied from contracts and negligence to civil rights and the First Amendment. Students were given page limits for their assignments, with none of them being permitted to write more than 20 double-spaced pages.

According to their responses to an email survey regarding teaching perspectives, legal writing instructors and professors at both schools shared many perspectives on teaching this year-end writing assignment. For example, many of these instructors/professors claimed that they had not discouraged students from using long quotations from cases (sometimes called "block quotes") and footnotes, but they also noted that most students had avoided frequent use of these rhetorical techniques. Generally speaking, citations in legal writing of this kind are in-line: all the relevant bibliographic information is included in a citation sentence or clause immediately after the name of the cited material or the assertion in the text that the cited material supports. The following is an example from paper 1019:[3]

> When a statute's plain language is ambiguous, a court may use legislative history to help determine Congress's intent. *See Safeco Ins. Co. of Am. v. Burr*, 551 U.S. 47 (2007). It is unnecessary to analyze the legislative history in this case because the text of § 101(2) is unambiguous and does not require a signed writing prior to the creation of a commissioned work. The legislative history does, however, provide further support for this conclusion. Committee Reports are the most authoritative source of legislative history. *Tellabs, Inc. v. Makor Issues & Rights, Ltd.*, 551 U.S. 308 (2007).

Note that in this example, the student cited two cases—*Safeco* and *Tellabs*—in citation "sentences" following the textual sentences that rely on the cited cases. Note, too, that there are no attributive cues in the textual sentences; conventionally, the assertion preceding a citation is attributed to the majority opinion in the case cited, unless certain special markers are used.[4]

Instructors/professors generally did not provide models for the types of briefs the students were to write. The textbook prescribed by each law school included one or two model briefs of the appropriate kind. Of ten instructors/professors who responded to an email interview about their teaching, only two supplied other examples, and both said they did so not to provide models of good brief-writing but rather to show what

[3] See Section 4.3.1 for a description of the conventions used in this study for numbering the papers of participants.

[4] Law-school-trained readers may note that "pincite" page numbers are missing from the citations in this example.

such briefs look like in practice. The students could use online research tools to find examples of briefs actually filed by lawyers in real cases, but the legal writing instructors did nothing to mediate students' assessments of the quality of such models, so it would have been difficult for students to select models, other than the textbook examples, upon which to base their own briefs. Nevertheless, students had been steeped for the better part of an academic year in reading court opinions; such documents are not written for the same purpose as memoranda, but students could be expected to model some of their linguistic practices on the opinions they had read.

As this section has shown, there is evidence the first-year law students at Academy School of Law and Lyceum Law College, though they no doubt varied a great deal in their personal characteristics and backgrounds, were all writing with very similar components of their cognitive environments accessible. Their training for the previous year prepared them with accessible assumptions about the typified situation of the memorandum and of legal writing in general. Their awareness of the importance of this assignment made their goal of success on it both accessible and strong. And their expectations of their instructors' expectations—the cognitive environments they imputed to their instructors—equipped them to adjust their writing styles to achieve their goals. Their year-end briefs are thus all of a single genre. This is true even across the law-school boundaries, owing to the similarities in the final assignments between the two schools and among the legal writing professors at Lyceum Law College. Of course, as I suggest in Section 7.2, it would be ideal to supplement the data in this study with qualitative interviews with the students to support (or undermine) this speculative evidence.

At least some of the conceptions of genre discussed in Section 3.5 might also call for the type of writing in question to be one that the writer engaged in repeatedly, the "conventional category of discourse based in *large-scale typification* of rhetorical action" described by Miller (1984, p. 163, emphasis mine). Or they may place the generic status of these texts in question because the classroom context makes the writing produced there "pseudotransactional" (Spinuzzi, 1996). Despite these concerns, students in these classes probably expected in the future to write texts in the genre or genres in which their assignments occurred. Their efforts to produce texts in a professional genre, even relating to hypothetical problems, likely constituted efforts with intense and accessible goals to evoke a reader response (here, from practicing attorneys acting as legal writing instructors/professors), based on the students' accessible (though perhaps weakly held) assumptions about instructors' expectations. Thus, though these students' efforts may fall short of satisfying the technical definition of genre espoused by some scholars, the students' work certainly represents a more motivated response to a shared rhetorical situation than any of the previous studies mentioned above.

4.2.2 Texts by single authors

As I noted in Section 3.6, researchers now often contemplate authorship as a collective and distributive activity. Blog posts are ghost-written. Twitter accounts are ghost-written and shared. Even published fiction is subject to concerns that editorial involvement in texts makes them collaboratively authored; the works of an author like Iris Murduch, whose resistance to editing makes them truly single-authored, are a rare exception (Pakhomov, Chacon, et al., 2011). It is my experience that in the professional context of law, court briefs often have many attorneys who claim authorship of them; a brief as filed might easily have four or five authors. Even the listed authors of a brief may not tell the story of authorship, given that associates in the law firm may be called on to draft segments of a brief edited, signed, and filed by a more senior lawyer.

Assessing gender differences in writing, however, demands that the texts studied be written by single authors, each of a gender recognized for purposes of the study. The writing assignments of first-year law students at Academy School of Law and Lyceum Law College address this concern because the schools limited students' ability to work together, and the structure of the assignments makes it unlikely that students will procure writing from outside.

In my view, the legal writing programs of both law schools in this study emphasize individual effort and assessment of the individual. Given the collaborative environment in which many professional legal briefs are written, this may seem strange. But law school is often an extraordinarily competitive environment; in the old days, it is reputed that students would intentionally misshelve books in the library to prevent their peers being able to use them for assignments (Turow, 2010). In fact, Academy School of Law's student honor code still expressly prohibited that practice in 2012. Legal employers are also acutely interested in students' class standing and individual level of achievement. So perhaps policies that prevent first-year students collaborating on writing and honor code provisions at both schools that assess harsh penalties for students working together are no surprise. They give rise to a much stronger presumption of single authorship than can be asserted with regard to the studies mentioned in Chapter 3.

Law students are also unlikely to be able to procure writing assignments from online banks of papers sold by other students (Ariely, 2012; Hansen, 2004). The law school writing assignments relate to complicated hypothetical problems, often involving case files with excerpts of evidentiary exhibits and testimony. No stock paper purchased online could ever hope to address the issues the students must take up in their writing assignments. Even if an instructor used a very similar hypothetical case from year to year, she need only make a slight change in the supporting materials to require the next year's students to take a much different tack in their analyses. Of course, it is possible that one law student could pay another or some third party to write her brief based on the current year's case materials. The amount of time required to do so makes it unlikely most law students could afford such a service; and the consequences for another law student to take on the task if she is caught make that unlikely, too. My own law students have occasionally told stories (always unsubstantiated, so far as I know) of other students who have parents or siblings who are lawyers who provide substantial editing services. Such circumstances would no doubt change the textual characteristics, but ghost-writing is a potential problem with any text not written before the researcher's eyes.

As this subsection has shown, the collection of samples from the first-year law students at Academy School of Law and Lyceum Law College resolves the single-author problem, at least to a reasonable degree of probability.

4.2.3 Authors who identify their own genders

A study of gender differences in writing ought to be very sensitive to the way it identifies writer genders. Section 3.4.1 showed that previous studies of gender differences in written communication suffered from limitations in this area. For example, some researchers relied unquestioningly on third-party assessments of author gender (Argamon, Koppel, Fine, & Shimoni, 2003; Koppel et al., 2002). Others used aspects of authors' computer-mediated communications to assess their genders and then used the resulting gender assessments to argue that aspects of the communicative performances varied with them (Rao et al., 2010). Still other studies had authors take gender-role assessment tests that raise serious concerns about gender stereotyping and a failure to address diachronic change in gender roles in American culture (Janssen & Murachver, 2004). These approaches might be described as "black-box," "question-begging," and "stereotyping" assignments of gender; this study avoids them by asking authors to identify their own genders.

I described the gender construct in this study in Section 3.4.2 (see p. 81) as a loosely and culturally defined set of social behaviors that are expected to make it possible to distinguish the two most common sexes from each other. I noted there that this study asked authors to identify their own genders.

But as Section 3.4 explained, even that approach poses problems because people generally do not have a sophisticated understanding of what *gender* means. They fill out surveys, questionnaires, medical forms and the like that ask them to specify their genders. Such instruments typically offer two choices, "male" and "female." But from some theoretical standpoints, it may be inappropriate to refer to these labels as gender labels as opposed to sex labels, while other theorists would oppose a bright line dividing sex and gender labels. And the average person, probably even the average law student, is not aware of these debates. A further problem arises if one considers transgendered persons. It is unclear where they are to check if given the option of two genders: male/female or masculine/feminine. For me, it is difficult to see how adding an "other" or "none of the above" option shows respect for research participants in my study.

The solution I chose for this study was to allow participants to identify their genders in a free-form questionnaire field in an online survey. In other words, students were asked their genders and allowed to write whatever they chose in response.[5] Of the 197 students who participated in this study, 193 responded to this question. Table 4.1 shows the results. (See Appendix E for the survey instrument and Section 4.3.1 for the procedure for administering it.)

As Table 4.1 shows, allowing for a free-form response creates a new problem: A proliferation of gender labels. Four different responses—*F, Fem, Female*, and *female*—came from participants who might describe themselves as being of a "female" or perhaps "feminine" gender. Four other responses—*Cis Male, M, Male*, and *Masculine*—came

[5] I'm grateful to Dr. Christina Haas for suggesting this elegant solution.

Table 4.1: **Self-reported genders of participants in present study ($n = 197$)**

Gender	Number	Percent of total
Cis Male	1	1%
F	5	3%
Fem	1	1%
Female	95	48%
female	3	2%
M	3	2%
Male	84	43%
Masculine	1	1%
Not answered	4	2%
Total	197	

Percentages rounded to nearest whole number, resulting in total tally of 103%.

from participants who conceivably consider themselves of a "male" or "masculine" gender.[6] Of course, while a researcher might presume that "F" was meant as "female," that may not be what the participant intended.

Rather than impose the associations of traditional gender identities on these participants, this study takes the approach of establishing an *ad hoc* research construct, in which authors may be assigned to "Gender F" or "Gender M." Authors who gender-self-identify with any designation beginning with the letter "F" (not case sensitive) are classified as Gender F. Those who self-identify with any designation beginning with the letter "M" (not case sensitive) are classified as Gender M. The prefix "cis" is ignored. Had there been any participants who used "tran" or "trans," they could have been classified as "Gender T."

As I argued in Section 3.4, all gender classifications are problematic and suspect. They are also subject to changing gender landscapes and expectations. Given that the common understanding of gender is that there are two (with possible accommodation for those who are transgendered or prefer to be ungendered) it is not unreasonable to

[6] The term "cismale" derives from gender studies, where it is used to refer to a person of the male sex who identifies with the masculine gender. Cisgendered persons thus contrast with transgendered persons in the congruity of their biological sex and the gender they feel or enact (DeFrancisco et al., 2014, p. 60).

group gender self-identifications based on two categories with similar linguistic features (namely their initial phonemes or graphemes). Though it comes with some challenges, it warrants greater credit than the gender-category assignments in the studies mentioned above.

In this section I have made the argument that the texts collected from students at Academy School of Law and Lyceum Law College are of a single professional genre and written by individual authors; and that 193 of them can reliably be labeled as being written by either Gender F or Gender M. The next section describes how these data were collected and prepared for analysis.

4.3 Data collection and preparation

Students at two law schools in the U.S. Midwest, referred to here with the pseudonyms Academy School of Law and Lyceum Law College, prepared a major writing assignment at the end of their first year in law school. The students' major project consisted of a memorandum of law or brief arguing for or against a dispositive motion before a hypothetical court in a hypothetical case. All the students in Academy School of Law wrote on the same hypothetical problem, set by the administrators of the legal writing program; at Lyceum Law College, individual section instructors developed their own hypothetical problems. I approached the directors of the legal writing programs at these two schools in the fall of 2011 and obtained their support for this research; they cleared it with their administrations.

After obtaining IRB approval for this study, I collected information regarding the structure of the course in which the writing samples were created by means of interviews with administrators of these programs and documents that they provided me.[7] This included information from the legal writing programs and instructors regarding the texts, assignment prompts, and model documents; as well as supplemental materials the instructors provided, whether they encouraged students to find and review examples of briefs of the kinds they were drafting, and the extent to which they emphasized various mechanical issues (grammar, citation, argument structure) in their instruction. That information provided valuable context that was described in Section 4.2.

[7] See Section 4.5 for details on the IRB process.

The use of these law-student texts allowed for the creation of a corpus of texts from a single professional genre, and the structure of the law school programs provided substantial confidence that the texts are single-authored. Section 4.3.1 describes the collection of these data and Section 4.3.2 the preparation they underwent before analysis.

4.3.1 Data Collection

This subsection describes the collection of the student papers, including a summary of the process for collecting data from students via an online survey and a brief description of the samples of writing collected. The legal writing program administrators of the two law schools cooperated in transmitting the invitation to participate in this study to the eligible law students at their schools. They arranged for me to provide them with the text of the invitation message, along with the *Information Sheet for Research*, a copy of which appears in Appendix D, and a link to the survey instrument, which appears in Appendix E. In spring semester 2012, the legal writing programs then sent the initial invitations about the time the final brief was due to be completed and followed up at weekly intervals for less than a month. Participating students were offered a $15 Amazon.com gift card for completing the survey and uploading their writing samples.

Survey instrument and data gathering

The survey instrument was developed according to the procedures outlined in Murphy (2002), using a process similar to that used by Eaton, Brewer, Portewig, and Davidson (2008) for an online survey. The survey instrument was hosted on Wufoo (http://wufoo.com), which permitted students to upload their writing samples at the beginning of the survey. The survey instrument is reproduced in Appendix E. It asked questions regarding student age, gender, highest previous degree, most recent writing course, and how the student learned English.[8] It also asked information about which section the student was in, so that this information might later relate the practices of particular teachers to peculiarities among their students's papers, if any.

[8] As Table F.4 shows, 14 participants had learned English as their second or subsequent language, and three learned English as their first language outside the U.S. These numbers were too small for me to assess whether the native-language status of students interacted with their genders in this study.

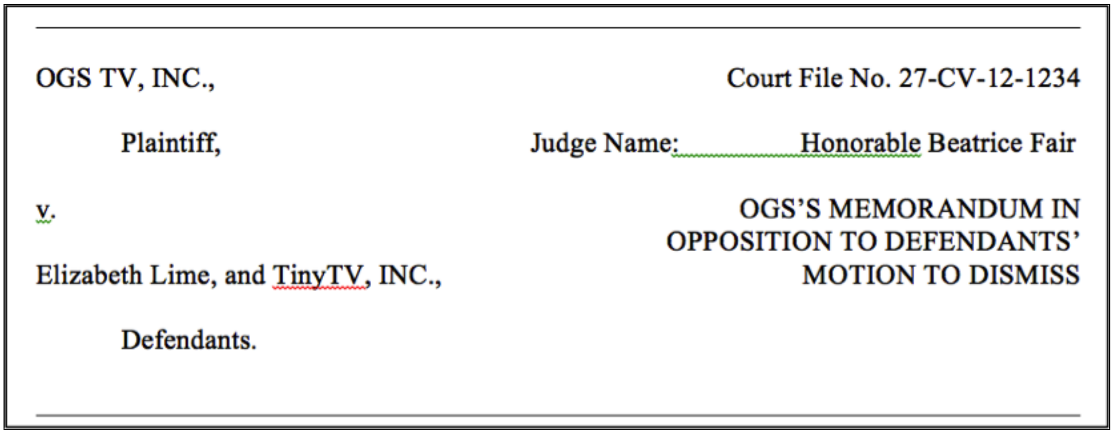
In all, 197 students completed the survey. According to law school records, 545 students were eligible; there was thus a response rate of approximately 36%. Though all questions on the survey were optional, 193 students provided information about their genders that could be interpreted according to the approach described in Section 4.2.3. Based on that approach and the responses, which are detailed in Table 4.1, the respondents were categorized into Gender F (n = 104) and Gender M (n = 89). (See Appendix F for a description of the respondents' other demographic characteristics.) The analyzed segments of student's briefs varied in length from 2,303 to 5,035 word tokens (including punctuation), with a mean length of 3,764 tokens. With one exception, all the papers were in Microsoft Word file formats; the exception, a PDF file, was converted into MS Word format using commercially available software. Before any other work on the briefs/memoranda, each file was reviewed to systematically remove all information that would identify the student author or the law school from the text of the file itself and from the file metadata.

All the writing samples and a spreadsheet of the survey responses are available on the Internet through the University of Minnesota Digital Conservancy. See Appendix A for a description of the files and their locations.[9] The student participants and their papers are designated in the data and throughout this dissertation according to a four-digit number assigned during anonymization. Paper numbers beginning with "1" originate with Academy School of Law and those beginning with "2" originate with Lyceum Law College.

Writing samples described

A preliminary assessment and review of the writing samples showed that the students had followed a largely formulaic approach to high-level structure similar to that suggested by the samples in the course textbooks. The memoranda were double-spaced, and each began with a caption of the kind shown in Figure 4.1 and concluded with a signature block like that shown in Figure 4.2. Some papers, including 1007, 1025, 1044, and 1098, had front or back matter that was not part of the memorandum itself. These elements included formal pleading documents like the motion, notice of motion,

[9] Generally, these materials will not be available until the dissertation is released in May 2017, or its results are disseminated in publications, or both.



OGS TV, INC.,	Court File No. 27-CV-12-1234
Plaintiff,	Judge Name: Honorable Beatrice Fair
v.	OGS'S MEMORANDUM IN OPPOSITION TO DEFENDANTS' MOTION TO DISMISS
Elizabeth Lime, and TinyTV, INC.,	
Defendants.	

Figure 4.1: **Caption from student brief**

and certificate of service. In the brief or memorandum itself, the structure was highly consistent:

- Caption: Every brief exhibited this.
- Introduction or summary: Not all briefs had this section (see papers 2084, 2091, 2093). In those that did, this section consisted of a brief introduction to the substance of the memorandum and the relief that student-attorney's client was seeking from the court. It may sometimes have been styled by the author as an "Issues" section (see papers 2026, 2095).
- Facts: Every brief exhibited this section, though it may have gone by other names, such as "factual background," "undisputed facts," and the like. In each memo, this section provided the facts of the instant case. This section was always the second longest section in the brief, after the argument section.
- Legal standard or standard of review: This section was not always present. If present, it articulated the standard for summary judgment or a motion to dismiss, the basis upon which the court would have to decide the motion. It was sometimes styled as "Procedure" (see papers 2057, 2086). Sometimes the content typical of this section appeared at the beginning of the argument section instead.

- Argument: This section appeared in every memo, and it was always the longest section. In this section, the student-attorney argued how the law, applied to the facts earlier discussed, should result in her client obtaining the relief requested of the court.

- Conclusion: The great majority of briefs included a section set off by a "Conclusion" header. Usually a paragraph or two at most, the conclusion reiterated the relief that the student-attorney was seeking from the court for her client and sometimes summarized the main points from the argument section.

Dated: March 15, 2012 **Respectfully submitted,**

____
Attorney for Defendants

Name
Attorney: #444454
Capital City, Moot 59321
(123) 678-9012
Email address

Figure 4.2: **Signature block from student brief**

Though students were discouraged by their textbooks from using footnotes in their briefs, some still chose to do so at least a few times (see papers 1035, 1043, 1070, all from Academy School of Law; footnotes were hardly present at all in papers from Lyceum Law College). Many students, however, used "block quotes," quotes of 50 words or more that conventionally must be indented on both sides and appear single spaced (*The Bluebook: A Uniform System of Citation*, 2011). Some used quite a lot of block quotes (see papers 1014, 1113, 2024, 2041). Figure 4.3 shows an example of such a quotation. All students used at least two levels of headings, one for the major sections identified above and one for key segments of their arguments. Some students used more levels of headings.

The argument sections of legal memoranda, at least at this level of law-student brief writing, typically conformed to a conventional argument structure, observed in the

The statute defines a work-made-for-hire as:

> a work specially ordered or commissioned for use as a contribution to a collective work, as a part of a motion picture or other audiovisual work, as a translation, as a supplementary work, as a compilation, as an instructional text, as a test, as answer material for a test, or as an atlas, if the parties expressly agree in a written instrument signed by them that the work shall be considered a work made for hire (emphasis added).

The statute requires that the work be 1) specially ordered or commissioned, 2) part of a collective work (such as a motion picture) and 3) that the parties expressly agree in a signed, written instrument. This case turns on whether an explicit oral agreement must be

Figure 4.3: **Block quote from student brief**

student samples. Legal writing textbooks usually tell students to construct paragraph level and higher levels of their argument using this structure—sometimes called "IRAC" (Clary & Lysaght, 2006; Schmedemann & Kunz, 2007), "CREAC" (Neumann & S. Simon, 2011), or "TREAT" (Murray & DeSanctis, 2009). Each stands for a roughly the same thing: IRAC for Issue, Rule, Analysis (or sometimes Application of the rule), Conclusion. CREAC has the student repeat the conclusion at beginning and end and adds "E" for Explanation of the rule. TREAT is the same as CREAC but offers "thesis" for "conclusion."

By contrast, there is little uniform guidance for writing fact sections. As a result, I speculated that the facts sections might be a part of the brief where the students wrote with less concern about the linguistic register and genre conventions of legal writing. In other words, the fact sections are places where gender differences in students' language might be more likely to endure, if indeed they were ever there to begin with. This study therefore examined the briefs both with and without the fact sections, and it examined the fact sections separately. However, as the findings of these three analyses were very similar, I have reported in this dissertation only the findings of the analysis of the full texts.

As a result of the processes described in this section, there were 193 texts written by law students at the end of their first year of law school, with each text classified by its author's self-identified gender. This sample of data is particularly useful for addressing the empirical research questions in this chapter. The plan was to assess the research questions above using statistics and machine learning and the features discussed in Koppel et al. (2002) and Argamon, Koppel, Fine, and Shimoni (2003). Before this could be done, the data required considerable pre-processing so that the texts could be described statistically and presented to the machine learning algorithms in abstracted form.

4.3.2 Data preparation: Annotating, splitting, tokenizing, tagging, and counting

Recall that the product of data collection was a corpus of 193 texts written by single authors writing in a disciplinary genre who had identified their own genders. Each text was thus labeled as having an author of Gender F or Gender M. The purpose of the data preparation described in this section is to produce tables of data—features or variables that describe the texts—for analysis using statistics and machine learning. This section provides a narrative summary of the efforts to prepare the data described in Section 4.3.1 for analysis. See Appendix G for a more detailed description of the activities described in this section.

The study described in this chapter was performed using statistics and machine learning algorithms and the lexical and quasi-syntactic text features used in Koppel et al. (2002) and Argamon, Koppel, Fine, and Shimoni (2003). Statistics provide an interpretable overview of the data and some gender differences in it. Machine learning has proven useful in earlier studies and permits many different analyses to be run on a large amount of data in a brief time. This study also functions as an example of the potential utility of machine learning for technical communication research. Other features could have been chosen for analysis, but the features in the Argamon/Koppel 02/03 study generally appear to be the largest set of features applicable to professional communication among the studies discussed in Chapter 3.[10] The Argamon/Koppel 02/03 study features consisted of the relative frequencies of:

[10] So, for example, X. Yan and L. Yan (2006) used emoticons as features, but these are absent from the samples in this study.

1. Approximately 400 function words. These include common words with low semantic content, such as determiners (*a, the, these, those*, etc.), prepositions, and forms of modal or auxilliary verbs (including *have* and *be*). This list also includes pronouns, words commonly used to structure relations between propositions (*although, consequently*), and common adverbs and adverbials (*probably, possibly, today, tomorrow*). In this study, I searched for 429 function words, but 84 of them did not appear in my corpus. Many words in the function-word list were archaic and not used in non-literary prose. For example, *thou* and *thy* appear on the function-word list but not in the corpus. Others simply did not appear. For example, *yourself* did not appear in the corpus at all. The complete list appears in Appendix C, and words not appearing in my corpus are noted.[11]

2. Parts of speech. Nouns (common and proper, singular and plural), verbs (present, past, participial, etc.), prepositions, etc. For purposes of this project, parts of speech include punctuation. I used the Penn Treebank tag set, and the complete list of part-of-speech "tags" appears in Appendix B.[12] Note that the tools used for this study did not distinguish between "count" and "non-count" or "mass" nouns.

3. The 100 most common part-of-speech bigrams in the corpus. Each bigram is a sequence of two parts of speech. The sentence [*The dog barked .*] includes three bigrams: Determiner–Common Singular Noun ("The dog"), Common Singular Noun–Verb (past tense) ("dog barked"), and Verb (past-tense)–Period ("barked . "). Examples of some of these bigrams in context appear in Appendix H, and the list of all the bigrams appears in Table I.4.

4. The 500 most common part-of-speech trigrams in the corpus. Each trigram is a sequence of three parts of speech. In the previous example, there are two: Determiner–Common Singular Noun–Verb (past tense) ("The dog barked"); and

[11] Argamon, Koppel, Fine, and Shimoni (2003) explained that the list of function words used for that study was available on the Internet as of the date of its publication. In 2013, I was unable to find the list at that address. After I contacted Professors Argamon and Koppel via email, Professor Koppel sent me a list of function words, but he expressed uncertainty as to whether it was indeed the list used in the article.

[12] I did not use four of the Penn Treebank tags; they are identified in the appendix.

> Common Singular Noun–Verb (past tense)–Period ("dog barked . "). Examples of some of these trigrams in context appear in Appendix H, and the list of all the bigrams appears in Table I.5.

See Section 3.3.1 for a fuller discussion of the meanings of these terms and illustrations of the bigrams and trigrams. In all, there were 986 such features, along with a few other features like total counts of tokens, bigrams, and trigrams in the texts. Chapter 7 suggests ways that this work might be extended with other features.

Before any other processing, the following text segments were excluded from analysis.[13] First, the formulaic caption and signature blocks described in Section 4.3.1 were removed, as they were nearly identical from paper to paper except for the hypothetical litigants' and attorneys' names. Second, legal citations were removed, as preliminary tests showed that they confounded the automated tools described below used to segment the texts. Third, headings were excluded, as their linguistic status was unclear. Finally, block quotes—the quotations of 50 words or more—were excluded, as they represent long stretches of text not composed by the students. No attempt was made to remove smaller quotations embedded within a student's text, however. So, for example, the following sentence appears in paper 1102:

> The general rule under the Copyright Act is that a "work protected under this title vests initially in the author or authors of the work."

Such a sentence represents a hybrid of the student's language and the language of the quoted text because the student integrates her original composition with that of the quoted text. I did not attempt to identify the instances where students used such quotations frequently or where they constituted a large percentage of the student's paper, which could prove to be a limitation.

At this point, all that remained was to calculate the relative frequency of each attribute for each paper. This involved splitting each brief-instance into sentences, breaking the sentences into word-tokens, and tagging those tokens for part of speech. Finally, the relative frequency of each feature for each paper or instance was calculated and the

[13] Note that though this section refers to the "deletion," "exclusion," or "removal" of some of the text segments, all the original texts are preserved in the data available online by the means described in Appendix A.

results were saved to a spreadsheet of feature values, with each row of the spreadsheet representing one paper and each column one feature. Based on the questionnaire data, a column was added with author gender information, with each of the 193 papers assigned either Gender F or Gender M.[14]

This section has explained the collection of data for this project, including the collection of questionnaire data identifying the gender of the author of each text and the selection of features to be analyzed in subsequent phases.

4.4 Data analysis

In its broadest terms, the goal of this dissertation is to explore the relation of gender and genre performances through the lens of cognitive pragmatic rhetorical (CPR) theory. One perspective on that issue is exploring whether gender variation in language is visible where genre is held constant. One way of exploring that question is with assistance of tools from statistics, natural language processing, and machine learning. In Chapter 3, I argued that other studies exploring this issue, whether they have used machine learning methods or not, have tended to operationalize the concepts or classes of *gender* and *genre* in ways that are not theoretically grounded. This study has sought to operationalize *gender* and *genre* in ways that illuminate the broader questions in this dissertation. I argued in Section 4.2 that collecting year-end capstone writing assignments from first-year law students would create a sample of texts all in the same professional genre and that gathering the student's gender self-identification would provide meaningful gender categories for those texts. Section 4.3.1 described the means by which these data were collected, and Section 4.3.2 described their preparation for analysis.

In this dissertation I have introduced CPR theory, and the empirical study described in this chapter is my attempt to illustrate CPR theory's utility in explaining the response of writers to a particular situation. Recall that the research questions for the empirical study described in this chapter are these:

1. Do Gender F and Gender M writers in a disciplinary genre in which they are being trained use lexical and quasi-syntactic stylistic features with relative frequencies that vary with their genders?

[14] See Appendix G.2 for details.

2. If so, do the differences appear in interpretable patterns?

3. Can machine-learning algorithms categorize the same texts by author gender based on the same features?

4. If so, do they provide interpretable models?

This section describes the analysis process for the empirical study.

The answers to the first two questions can best be obtained by statistics, which permit the researcher to conclude whether the difference in relative frequency of a feature between Gender M and Gender F authors is statistically significant. As a preliminary matter, note that there is always some difference between the values. For example, no two authors of papers used common nouns with exactly the same frequency. Consequently, comparison of a feature in all the papers of authors of one gender with those by authors of the other gender will also always result in at least some difference. Statistics answers the question of whether that difference is likely to be meaningful or just the operation of chance. Section 4.4.1 describes the tests used in some detail.

The answers to the third and fourth research questions require the application of machine learning algorithms. Section 4.4.2 offers a justification for the application of MLAs and an overview of the process.

4.4.1 Statistics

One objective of this study is to be able to assess whether the values of features in each corpus under study vary depending on author gender. This approach is taken in part in response to Argamon, Koppel, Fine, and Shimoni (2003) and their reliance on the informative/involved dimension described by Biber (1995). All the features described above and collected for this study were examined for the presence of statistically significant differences between the texts written by Gender F and Gender M authors. This study also examined aggregated features to situate its findings alongside those of the earlier studies by Argamon and his colleagues and by Biber.

The variation between the values for Gender F and Gender M authors has to be assessed with a measure of statistical significance. For example, say the mean value for the relative frequency of the common noun (singular) part of speech in the corpus for

Gender M authors is 0.1734 (SD 0.0135) and that for Gender F authors is 0.1742 (SD 0.0145). Given that none of the instances in the samples has the same value as any other instance for this feature, we would expect some difference between the Gender M mean and the Gender F mean. It would be an extraordinary coincidence if they were identical. But what kind of coincidence would be necessary for them to be just these values? In other words, what is the chance that these two means vary by this amount?

For two independent samples of data, one for the Gender M authors and one for Gender F, I wished to know whether the difference between them appearing in my sample reflects a true difference in the population or is merely the result of sampling. In statistical terms, the *null hypothesis* is that the two samples *are not* significantly different, and the *alternative hypothesis* is that they *are* significantly different. A common statistical method for resolving this question is the *two-sample* t-*test* (Utts & Heckard, 2006, p. 450). In that case, the result is a p-value that indicates the strength of the evidence for rejecting the null hypothesis. The lower the p-value, the greater the evidence for rejecting the null hypothesis and accepting the alternative hypothesis, that is, that the values in the two samples really are different. How low the p-value must be before rejecting the null hypothesis is referred to as α and is a matter of disciplinary convention, but the most common α is 0.05 (Utts & Heckard, 2006, p. 448). That is the value used in this study.

One condition for use of the two-sample t-test is that feature values represent a Gaussian distribution for each sample.[15] Unfortunately, in the data for the features in this corpus, a test for Gaussian distribution revealed that fewer than one in seven of the features may have a Gaussian distribution. Unless some other recognized distribution or parameter of the feature values is identified, they are regarded as having a non-parametric distribution. Consequently, a different test, called the *Mann-Whitney test* (which is equivalent to the *Wilcoxon rank-sum test*), is appropriate (Utts & Heckard, 2006, p. 614). Once again, this study used an a priori α of 0.05 for the significance

[15] Often called a "normal" distribution, but I worry about the implications of the term "normal" in other contexts and try to avoid it in this one.

threshold for Mann-Whitney.[16] Reporting a difference in frequency where the *p*-value is greater than or equal to α arguably represents a misunderstanding of the data. Thus, in the example given above, if the *p*-value were 0.779, and α were 0.05, it would be inappropriate to say, "Gender F authors use common nouns more frequently than Gender M authors." Such a claim would be appropriate only where the *p*-value is less than α. Nevertheless, for the presentation in Chapter 5, it will be helpful to be able to refer to differences as "significant" (representing a differences where the *p*-value is less than 0.05) and "non-significant" (where the *p*-value is 0.05 or greater).

As I note in Section 7.4, I would like to do effect size calculations on my data, but I have not so far done so because of the difficulty of selecting a method for calculating effect sizes on samples that are not in Gaussian distributions.

Chapter 5 summarizes the results of the application of the Mann-Whitney test to the means for Gender F and Gender M of each of the 986 features described above. In addition, the mean relative frequencies for certain other categories were calculated because they were used in the statistical analyses in Argamon, Koppel, Fine, and Shimoni (2003). These calculated values summed the relative frequencies of various function words (calculated as described above):

- First-person singular pronouns: *I, me, my, mine, myself.*
- First-person plural pronouns: *We, us, our, ours, ourselves.*
- All first-person pronouns: The sum of the previous two values.
- Second-person pronouns: *You, your, yours, yourself.*
- Third-person singular feminine pronouns: *She, her, hers, herself.*
- Third-person singular masculine pronouns: *He, him, his, himself.*

[16] There is an argument for using a much lower value for α here. Utts and Heckard (2006) discussed the multiple-comparison problem: when one compares two samples using multiple variables, the probability that one variable will vary between the samples merely by chance increases as a statistical matter. There are methods for adjusting the *p*-value to address this concern, including the Bonferroni adjustment, but I have not employed them here. The failure to do so poses no threat to my findings and discussion of them, because lowering the *p*-value means that fewer features would be statistically different between Gender F and Gender M authors—supporting my finding below that there is no pattern of stylistic difference between these two groups in these samples.

- Third-person singular pronouns: The sum of the previous two values.
- Third-person plural pronouns: *They, them, their, theirs, themselves.*
- Third-person pronouns: The sum of the previous two values.
- Contractions: Including all instances of *n't, 'ld, 've*, etc.[17]

It and *its* were assessed separately from the other third-person pronouns as impersonal pronouns.[18] Some of these features are not present in this corpus: For example, *yours* and *yourself* do not appear in any of the papers studied.

The findings illustrate that a number of features that were used more or less frequently by Gender F in Argamon, Koppel, Fine, and Shimoni (2003) show no significant difference in this study; but there were some features where significant differences appeared.

4.4.2 Machine learning

I performed machine-learning analyses of the data in this study, 15 trials with all 986 features that I gathered and 17 trials with reduced feature sets produced in the fashion described below. In this subsection I offer an argument for the application of machine learning to the empirical study in this dissertation and explain its operation in that context. This includes a description of the inputs and outputs and of machine-learning algorithms in general, how the results are evaluated, and the specific MLAs applied in this study. Justifying the use of machine learning algorithms in theoretical research requires first explaining what *machine learning* is and second explaining how it can contribute meaningfully to our knowledge. Section 3.3.1 (beginning at page 64) introduced MLAs. For purposes of the study described in this dissertation, machine learning is the use of computer algorithms to identify patterns—possibly very subtle ones—in texts to determine whether the texts can be classified based on lexical, quasi-syntactic, or other

[17] Note that the NLP tools used here decouple the contraction from the word to which it is attached. Thus, "he's" becomes *he* and *'s*, and "don't" becomes *do* and *n't*.

[18] As far as I can tell, these tools do not distinguish between the "expletive" (or "ambient" or "non-referential") *it*—which serves a syntactic role but has not semantic value—and pronominal *it*. For an example of expletive *it*, consider: "It is pleasant to work outdoors in the spring." For pronominal *it*, consider: "I ate it." In the latter, *it* stands in for some noun already in focus in the conversation.

textual characteristics so that the classifications are consistent with a category variable, such as author gender.

Machine learning has received little to no attention in the scholarly literature of disciplinary, professional and technical communication (DP&TC).[19] A search in fall 2014 of the indices of the journals *Technical Communication Quarterly*, *Technical Communication*, and *Journal of Business and Technical Communication* revealed no studies examining machine learning or its applications. Even a search for relevant literature in the proceedings of the Special Interest Group on Design of Communication of the Association of Computing Machinery (ACM SIGDOC)—a technical communication group that is not averse to technical topics—reveals only two essays, each of which makes reference to machine learning without actually using it (Kelly, Abbott, Harris, DiMarco, & Cheriton, 2010; Kelly, McDougall, & Abbott, 2009).

Nevertheless, machine learning has figured in research calculated to result in tools for achieving organizational goals in the medical, legal, and finance fields, all of which are normally objects of interest for researchers in technical and professional communication. For example, three studies conducted in the field of "medical informatics" exemplify the ways in which machine learning is coming to be used in that field. The medical industry generates a great quantity of data about patients and treatments, much of it structured into codes based on industry standards like the federal government's ICD-9 diagnosis and procedure codes and psychologists' *Diagnostic and Statistical Manual of Mental Disorders* (Centers for Medicare and Medicaid Services, 2014; American Psychiatric Association, 2013). Computer algorithms, including machine learning algorithms, can be applied to such codes fairly easily, as they constitute distinct features or categories that can automatically be identified from medical *texts*. However, there remain in many medical records *free form* data fields consisting of physicians' notes or transcribed dictation from physicians. Such "unstructured text" (Pakhomov, Hanson, et al., 2008, p 198) can be assessed with a combination of the machine learning and NLP techniques described in this chapter. Pakhomov, Hanson, et al. (2008) used machine learning to examine clinical notes sections from examinations of 145 patients with diabetes, looking for evidence that the mandatory three-component diabetic foot examination had been given. Pakhomov, Shah, et al. (2010) conducted a study using the records of

[19] See page 2 for a definition and discussion of this term.

515 medical patients to determine which were using aspirin and whether their records evidenced contraindications for aspirin use (such as allergy to aspirin, etc.). McCart et al. (2013) similarly used machine learning on unstructured text in hospital records to identify cases where patients had fallen.

In the legal context, Ashley and Bridewell (2010) introduced a special issue of the journal *Artificial Intelligence and Law*, in which several articles addressed machine learning techniques, by explaining the application of machine learning (among other techniques) to the problem of retrieving relevant documents from party records during the discovery phase of lawsuits. During this phase of the lawsuit, each party demands from the other documents that are relevant to certain questions at issue in the litigation; the objective is to obtain information that will either be admissible at trial or will lead to information admissible at trial. Application of automated tools to electronically stored documents is increasingly referred to as e-discovery (p. 312). Given the volume of documents produced during discovery in complex litigation, the legal industry stands to benefit greatly from the application of automated tools to identify useful documents. Finally, in the field of finance, Humpherys et al. (2011) examined the unstructured textual portions of reports filed with the SEC—the managers' narratives surrounding the financial data—and calculated as features the relative presence of a variety of linguistic cues to see if a machine learning algorithm could successfully detect a filing from a fraudulent company based on the text.

The introduction of machine learning in this study thus serves two purposes: First, it explores the use of machine learning algorithms as a method of analyzing data. Consequently, RQ3 asked whether machine-learning algorithms could categorize texts by author gender based on selected features. Second, it takes a first step toward making machine learning itself an object of study in DP&TC. Thus, in the event the answer to RQ3 is "yes," RQ4 asked whether the models, or knowledge representations, created by MLAs are interpretable. On these grounds, the use of MLAs for this project is justified.

As noted in Section 3.3.1, MLAs take as their inputs *features* about *instances*. The instances are the units of study. The data collection and preparation efforts described in Section 4.3.2 resulted in a corpus representing a portion of the memoranda or legal briefs submitted by 193 students at the end of their first year of law school. Each student's paper in the corpus is therefore an instance for the application of machine learning. The features, sometimes called *attributes*, are variables that describe or function to create an abstraction of the instances. (See Figure 3.2 at page 65 and the accompanying text for a more detailed explanation and example.) As with any investigation, the researcher must select the attributes or variables to be considered. This study used attributes based closely on those in the Argamon/Koppel 02/03 study (Argamon, Koppel, Fine, & Shimoni, 2003; Koppel et al., 2002), described more fully in Section 3.3. Using this large set of features provides a greater opportunity to identify differences that may constitute an interpretable pattern. In the present study, the corpus was abstracted into 986 lexical and quasi-syntactic features or attributes, basically relative frequencies of certain function words, parts of speech, and patterns of parts of speech (bigrams and trigrams), for each of 193 instances (the 193 students' texts). So for each instance or memorandum, hundreds of features measure the relative frequency of various function words, parts of speech, and part-of-speech bigrams and trigrams. These features represent to a certain extent the authors' stylistic choices; because they consist of lexical and quasi-syntactic features we would expect not to be content- or topic-dependent.

The outputs of machine learning algorithms are concept descriptions, what Witten et al. (2011) also call "knowledge representations" and describe as "descriptions of the structural patterns in the data"(Witten et al., 2011, p. 61). The *concept* to be described in this study is author gender, and because this study explores a *classification* problem—whether the texts can be classified according to author gender—the gender of authors is also the *class* of the instances. The form of the concept description depends on the input data and the machine learning method applied to it. Section 3.3.1 provided accounts of several types of machine learning algorithm. Another way of thinking of the output or concept description of an MLA is as a heuristic, a basis for deciding how to classify instances without knowledge of their category labels.

The data about instances upon which the algorithm bases its heuristic are usually referred to as the *training set*, and the data about instances against which the heuristic's effectiveness for prediction or classification are tested are usually referred to as the *test set* (Jurafsky & Martin, 2009, pp. 91-92). Often, a data set will be divided into disjoint *folds* to permit all the data to be used both for training and testing, a process called *cross-validation* (Jurafsky & Martin, 2009, pp. 154-55). Assuming 10 folds, the instances are divided into 10 batches and the researcher's tests are run 10 times, once with each batch as the test set and the other nine combined as training set. The error rates for all ten runs are averaged to calculate a mean error rate. Cross validation permits the researcher to use all the data in the corpus as training and test data, and it also helps to prevent the resulting model from being *overfitted* to the training data (Hawkins, 2004). *Overfitting* occurs where "the learning algorithm adapts so well to the given data, that noise or particularities of the specific sample are also encoded by the learned model" (Wu & Shapiro, 2006, p. 433). Overfitting may create models that are good for classifying the instances used to create the model, but they may do a poor job of predicting the classification of previously unseen instances (Gigerenzer & Brighton, 2009).

The MLAs applied in this study are those implemented in the WEKA machine learning framework (Hall et al., 2009; "Weka 3: Data Mining Software in Java," n.d.; Witten et al., 2011). WEKA is open-source software available free of charge from the University of Waikato in New Zealand. Because the authors of Witten et al. (2011) are principally responsible for the development and maintenance of WEKA, this dissertation relies on their work for a description of the WEKA tools and for guidance on how to implement them. The corpus described above was subjected to classification by the following MLAs in WEKA, with each application of an MLA to the corpus described here as a "trial."

- `Winnow` (and balanced `Winnow`): These algorithms generate linear models and are included here principally because Koppel et al. (2002) used balanced `Winnow`.

- Other linear models: `Logistic`, `SimpleLogistic`, and `VotedPerceptron`.

- Instance-based classifiers: `IB1` and `IBk`. In the strict sense, these classifiers do not build models. Rather, they store the training set and compare each instance in the test set to the instances in the training set at the time of classification.

- Support-vector machines: `SMO` and `SPegasos`. These algorithms combine aspects of instance-based and linear models.

- Bayesian: `NaiveBayes` (with and without kernel estimator).[20] These algorithms

[20] The kernel estimator is an alternative to basic `NaiveBays`, advised by Witten et al. (2011) for feature sets where continuous variables exhibit a non-parametric distribution. See the discussion in Section 4.4.1 regarding the lack of Gaussian distribution for most features in this study.

take advantage of Bayes' theorem to generate a concept description in the form of a *Bayesian belief network*.

Thus, there were 15 trials with the full 986-feature set. For each MLA, Appendix J includes a brief description of its method of functioning; the attributes with which it can function; necessary transformations of the data sets, if any, to implement the MLA; identification of parameters used with the MLA in the trials; and detailed findings from all the trials. Chapter 5 summarizes these findings. Each MLA was applied to the instances in the corpus using all 986 features described in Section 4.3.2, except those features that were not used by any of the participants.

Despite the fact that machine learning algorithms (MLAs) "are designed to learn which are the most appropriate attributes to use for making their decisions.... adding irrelevant or distracting attributes to a dataset often confuses machine learning systems" (Witten et al., 2011, p. 307). Attributes that are statistically related to each other or redundant can also cause problems. A solution to these problems is *attribute selection*, or the selection of a subset of available attributes in the dataset. This study included the use of attribute selection to optimize the performance of the MLAs in trials. A brief description of the process of attribute selection is therefore appropriate.

There are two choices for attribute selection: "scheme-independent" and "scheme-specific" selection. The former uses machine learning or statistical analysis to identify only those attributes that are relevant to classification and that have little intercorrelation among themselves. The latter approach, and the approach used in this study, is tailored to the specific MLA that will be applied to the data: it determines which attributes will maximize the success of that particular MLA. Attribute selection can improve MLA performance, but it offers a second advantage: it identifies a subset of features in the data that are most useful for distinguishing between texts written by authors who self-identify as Gender M and those who self-identify as Gender F. These subsets of attributes may help to make the machine learning models intelligible, even when the models they create might otherwise not appear intelligible.

For each of the MLAs above, one trial was run with the full 986-feature data set and another was run with a dataset created especially for that MLA by WEKA. WEKA's `ClassifierSubsetEval` module with default configuration, including the `BestFirst` method for attribute selection, was used for attribute selection, except that Appendix J.5

describes a second round of attribute selection for the `NaiveBayes` classifiers. Thus, there were 32 trials in total: 15 with the MLAs described above using the full feature set, 15 with the MLAs described above with reduced features sets selected by `ClassifierSubsetEval`, and two with the `NaiveBayes` classifiers with reduced feature sets selected by `WrapperSubsetEval`

One other matter deserves mention here: The identification of significant MLA classifier results; in other words, how accurate a machine-learning classifier has to be for the results to be interesting. There are two measures, both of which are applied to trials in this study: statistical significance and practical significance. Statistical significance assessments determine whether the machine learning algorithms (MLAs) could achieve some measure of success merely by randomly assigning texts to gender categories. Practical significance is another matter: It raises the question whether the results, even if statistically significant, are of any practical interest. In short, the results are practically significant only if we can offer an interesting answer to the question, "So what?"

Assessing statistical significance of results entails comparing an MLA's assignments or predictions of the gender of a text author to the actual self-reported gender. This *observed agreement* is calculated in a straightforward way: For example, if a machine learning classifier is to classify 193 texts, each into one of two categories (A or B), then the observed agreement is the number of instances or texts correctly classified divided by the total number of instances in the corpus, in this example, 193.

Merely calculating observed agreement, however, does not account for the underlying distribution. If in the previous example, 180 of the instances are in Class A and only 13 in Class B, an automated classifier can achieve observed agreement of 93.3% or 0.933 merely by classifying every text in Class A. Most of the agreement in that case would be entirely due to chance. A common way of addressing this issue is the *most frequent class baseline* (Jurafsky & Martin, 2009, p. 155). This approach compares observed agreement of a trial with the agreement that would be achieved merely by assigning the most frequent class to all instances. In the case of the empirical study in this dissertation, 104 of the 193 samples have Gender F authors, according to authors' questionnaire responses. Thus, assigning Gender F as the author gender of all 193 texts results in correct classification 53.87% of the time ($104/197 = 0.5387$). So for the results of an MLA trial to be statistically significant, the MLA must classify test instances correctly more than 53.87% of the time.

But even here, a higher observed agreement could be the result of chance. One time out of ten, or one time out of a hundred or a thousand, an MLA is expected to classify instances correctly a high percentage of the time just based on chance. It is perhaps unlikely, just as it is unlikely that rolling five dice together will result in five sixes. As with the dice, though, it is possible to calculate the probability that a given result arises randomly. A paired t-test allows the researcher to determine whether any increase (or decrease) in agreement from the most frequent class baseline is statistically significant. The result is a p-value that expresses the probability that the results arise from chance. The choice of how much risk of chance variation is tolerable is up to the researcher and is usually constrained by disciplinary conventions: in the social sciences, for example, a p-value of less than 0.05 is often accepted as a signal of statistical significance (Utts & Heckard, 2006). The threshold for p-value established before a study is referred to as α. For this study, then, α=0.05 is the boundary for statistical significance.

In the WEKA machine learning framework, the `ZeroR` algorithm classifies all the test instances in the most frequent class (Witten et al., 2011, p. 459). Its level of observed agreement is therefore equal to the most frequent class baseline. The WEKA framework is also capable in its experiment mode of comparing the results of the `ZeroR` algorithm/most frequent class baseline to the results of another MLA using a paired t-test. Thus, the trial results in Appendix J are always presented with the left-most column showing the `ZeroR` baseline and the other columns showing the performance of the MLA(s) being tried. Results where the MLA performed statistically better or worse than `ZeroR` are marked in the applicable tables. As Chapter 5 shows, MLAs performed statistically better than `ZeroR` in 9 of the 32 trials I ran.

Practical significance is a more subjective call. I chose to consider the results of a trial practically significant if it classified the texts in this study accurately (observed agreement) at least 66% of the time. The choice of the 66% level for observed agreement does not arise from a mathematical calculation or statistical test; it is rather my judgment, reached by the following reasoning. Chapter 3 and Section 4.4.2 identified several studies that have used machine learning algorithms for classification problems. A review of those studies and of related literature revealed that each published study had at least one trial with at least 66% classification accuracy. Several studies have been published reporting observed agreement in the range of 66% (for example, Humpherys et al. (2011)) to 72% (for example, Rao et al. (2010) and X. Yan and L. Yan (2006)). Based on the willingness of other researchers to deem a study's results worthy of publication and report, we can conclude that this level of observed agreement is of at least moderate interest to other scholars and that lower levels of classifier accuracy have not been sufficiently interesting to the researchers who obtained them or to peer reviewers to result in publication. As Chapter 5 shows, this choice of boundary for practical significance results in less than half—4—of the 9 statistically significant trials being deemed practically insignificant.

In summary, machine learning algorithms provide the means to identify patterns in data that correspond to variables of interest, in this case, authors' self-reported genders. As inputs they take abstractions of the text instances in the form of numeric attributes indicating the relative frequency with which authors used certain word types, parts of speech, etc. Based on the instances, the MLAs build models that allow them to take advantage of large numbers of features to identify subtle patterns in the texts and use them for classification. Their performance can be evaluated for statistical significance based on the most frequent class baseline and for practical significance against the performance of classifiers in earlier published studies. Before moving to the findings of these efforts in Chapter 5, the next section addresses ethical concerns in the design of this study.

4.5 Ethical considerations

The design of this empirical study addressed a broad array of ethical issues. Important among them are issues of participant consent, anonymity of the reporting, institutional review board compliance, permission from participants to reproduce texts, reliability and validity, and a need for continual reflection on ethical issues by researchers (Breuch et al., 2002). This section addresses some of those efforts, including the design of the participant consent form and its approval by an institutional review board; copyright law issues associated with the use of students' texts; and efforts to make the data of this empirical study available to other researchers.

The consent form for this study was based upon contents suggested in Breuch et al. (2002, p. 11), to meet the requirements of the University of Minnesota's Institutional Review Board, which were set out in its Protecting Human Subjects Guide (Board, 2004, p. 5) and on its web site. See Appendix D for the complete consent form. The University of Minnesota IRB/Human Subjects Committee determined that this study is exempt from review under federal guidelines 45 CFR Part 46.101(b), Category #2 "surveys/interviews; standardized educational tests; observation of public behavior" (Study Number: 1202E10685).

Because this study involves the copying and transformation of texts by students that they have fixed in a tangible medium of expression, the sample texts are subject to U.S. Copyright law, Title 17 of the United States Code. My intention is to publish the texts on the University of Minnesota University Digital Conservancy so that other researchers can reproduce this study's efforts using the same texts. The consent form included a grant of a license from each student for those purposes for this study and for any similar studies conducted by researchers using the same texts.

Questions of reliability and validity are important. Irresponsible claims based on incomplete data or poorly designed studies can cause real harm to research participants and other people like them who have never consented to be part of the research. I hope this study design has gone some distance toward addressing these concerns, but I have taken the further step of publishing almost all of the data and materials created for this study on the Internet. Appendix A describes in detail the information deposited with the University of Minnesota's University Digital Conservancy. This should permit other investigators to replicate the results of this study with the same data or to analyze the work for flaws. The choice of the Digital Conservancy also reflects a concern about the durability of storage sites for some research materials. As noted in the footnote at page 120 in Section 4.3.2, the list of function words used in the Argamon/Koppel 02/03 study—published in 2002 and 2003—was not available in 2013 at the URL identified in the associated articles—a casualty of "link rot." Clearly, we should not look only to researchers—even ones like Professors Argamon and Koppel, who continue to publish actively—to maintain long-term access to materials associated with their published works. The Digital Conservancy seems an obvious solution because it promises to make the materials present there available in perpetuity.

Finally, this study takes to heart the admonition in Breuch et al. (2002) and elsewhere to consider ethical issues throughout the research process. I have kept and will continue to keep a journal of participant interactions and research project decisions to guide work; some of the journal reflections may be appropriate for discussion in the published research from this study, if there is any, and if the journal entries may help readers to understand the ethical context of the study and my decisions.

4.6 Conclusion

This chapter has described the design of the empirical study in this dissertation, which contributes to answering following specific research questions:

1. Do Gender F and Gender M writers in a disciplinary genre in which they are being trained use lexical and quasi-syntactic stylistic features with relative frequencies that vary in relation to their genders?
2. If so, do the differences appear in interpretable patterns?
3. Can machine-learning algorithms categorize the same texts by author gender based on the same features?
4. If so, do they provide interpretable models?

I will use cognitive pragmatic rhetorical (CPR) theory in Chapter 6 to explain the findings regarding these empirical research questions. The overarching research question presented by this dissertation is how CPR theory can help us understand disciplinary, professional, and technical communication (DP&TC).[21] As I explain in Chapter 6, CPR theory explains how training into the conventions of a disciplinary genre could overcome the gendered linguistic habits of the trainees to make their writing indistinguishable, or nearly so, based on features previously shown to be useful for distinguishing texts written by males from texts written by females. The answers are significant for our understanding of the development of genre knowledge in apprentice members of disciplinary communities and for addressing concerns that other studies of gender difference

[21] See page 2 for a definition and discussion of this term.

in language have essentialized differences between males and females that may prove to be malleable.

I argued in Section 4.2 that the law-school context provides a unique opportunity to answer these questions by providing a sample of texts in the same genre, written by single authors whose genders are known. First-year law students write a year-end capstone writing project, usually in the form of a motion-practice brief. The students themselves are subject to powerful incentives to conform to the disciplinary conventions of genres to which law school exposes them. Section 4.2.1 concluded the briefs that students from two law schools, styled here as Academy School of Law and Lyceum Law College, prepared in spring 2012 represent their written responses to strikingly similar situations, prepared for strikingly similar purposes. In the terms of Chapter 3, they represent texts in the same *genre*. Second, in Section 4.2.2, I argued that these texts are more likely to represent the efforts of single authors than most of the previous studies discussed. Third, the authors of these papers identified their own genders. Section 4.2.3 showed that grouping their responses according to linguistic similarity into two categories, Gender F and Gender M, was justified.

Section 4.3.1 described the collection of the writing samples and related information and offered a formal description of the samples. There were 193 papers for which I collected gender information. Section 4.3.2 offered justification for excluding some portions of the samples from analysis.

In Section 4.3.2, I proposed that the features to be used for this study should be those used for the Argamon/Koppel 02/03 study, because that study has proved influential among other studies described in Chapter 3. The features consist of relative frequencies of function words, parts of speech, and certain part-of-speech bigrams and trigrams. Section 4.3.2 also explained the process for calculating each feature for each paper. The result was a data set, which can be conceived of as a spreadsheet where the rows represent the instances or student papers, and the columns represent the features.

Section 4.4 offered a rationale for performing statistical tests on the resulting features and for applying machine-learning algorithms to them. This combination of analytical methods increases the likelihood that at least some of the results will be interpretable, because statistics permit the researcher to conclude which features vary based on author gender according to a threshold of statistical significance adopted for this study. At the same time, MLAs open the possibility of text classification by author gender with the use of features that may not be significantly different from a statistical perspective, but may still represent subtle patterns in the data. Even with the MLA trials, though, Section 4.4.2 explained that tests of statistical and practical significance are appropriate.

This chapter concluded with discussion of ethical issues in Section 4.5. Among other things, that section described to decision to deposit materials from this study in the University of Minnesota's University Digital Conservancy. The materials there are described in Appendix A.

The context for data collection for this study, the collection and preparation of the data, and the analysis of them described in this chapter offer the hope of answering the research questions here. Chapter 5 gives a summary of the findings from these activities, which show that these authors did not exhibit the differences previously attributed to male and female authors. But the findings also show that some differences appeared. Chapter 6 explains how CPR theory accounts for these findings.

Chapter 5

Findings: Gender similarity in genred writing

5.1 Introduction

This chapter reports findings of the statistical analysis and machine learning algorithm (MLA) trials used for the empirical study in this dissertation. The data collection and preparation efforts described in Chapter 4 resulted in a corpus containing memoranda or legal briefs submitted by 193 students at the end of their first year of law school, with some formulaic portions of the texts excluded from the analysis. This corpus was abstracted into 986 lexical and quasi-syntactic features or attributes—basically relative frequencies of certain function words, parts of speech, and patterns of parts of speech (bigrams and trigrams)—for each of the 193 instances.

The overarching inquiry in this dissertation is whether cognitive pragmatic rhetorical (CPR) theory can contribute to our understanding of rhetorical and disciplinary, professional, and technical communication theory and in particular to our theories of gender and genre performances. In support of that inquiry, I have conducted an empirical study, with the following research questions:

1. Do Gender F and Gender M writers in a disciplinary genre in which they are being

trained use lexical and quasi-syntactic stylistic features with relative frequencies that vary with their genders?[1]

2. If so, do the differences appear in interpretable patterns?

3. Can machine-learning algorithms (MLAs) categorize the same texts by author gender based on the same features?

4. If so, do they provide interpretable models?

This chapter provides detailed answers to the research questions, reporting the findings from applying two methods of analysis to the instances and their attributes: statistical analyses and machine learning analyses. First, the statistical analyses explored whether the differences in relative frequencies of the attributes in each instance differed by gender on the "involved" and "informational" dimension first described by Biber (1995) and used by Argamon, Koppel, Fine, and Shimoni (2003) to explain the results of their study; this study then considered whether any pattern of other statistically significant features appeared in the data. The machine learning trials showed whether MLAs can classify these instances—the students' papers—according to authors' self-identified genders using the attributes. See Sections 3.3.1 and 4.4.2 for a fuller discussion of the operation of MLAs.

As this chapter shows, of 28 features previously associated with the "involved" and "informational" dimension by Argamon, Koppel, Fine, and Shimoni (2003), Gender F and Gender M writers in this study used only three with frequencies that differed significantly, and in two of those cases, the present study's significant results conflicted with the earlier study. Writers in this study *did*, however, use certain other lexical and quasi-syntactic stylistic features with relative frequencies that varied significantly with their genders. For example, Gender F authors used certain time and rhetorical sequencing adverbials, like *furthermore*, more often than Gender M authors. These differences provide little in the way of interpretable patterns, however, because there were many other similar features where there were no significant differences and where the non-significant differences "cut the other way." For example, for many time and rhetorical sequencing adverbials, such as *accordingly, already, although, before, lastly,*

[1] Sections 3.4 and 4.2.3 provided an account of the rationale for classifying the authors in this study as "Gender F" and "Gender M" based on participants' gender self-identification.

later, moreover, and *soon*, there was no significant difference between the Gender M and Gender F authors. And for some among these examples, including *later, moreover*, and *soon*, Gender M writers used the word more frequently on average, though not significantly so. These findings are described in detail in Section 5.2.

Second, this chapter shows that machine-learning algorithms were generally not able to classify these 193 texts by author gender at levels defined in Chapter 4 as *statistically* or *practically* significant when they were trained using the full set of features. Nevertheless, three MLAs—`NaiveBayes`, a Bayesian MLA, `SMO`, a support vector machine, and `SimpleLogistic`, an algorithm that learns a logistic regression model—*were* able to classify the texts in this corpus at levels that achieved statistical and practical significance after the full feature set in this study had been reduced to maximize the performance of each MLA with each corpus.

The best performance was by `NaiveBayes`, which achieved observed agreement of 73.19%. This performance was arguably similar to that in Koppel et al. (2002),[2] which achieved observed agreement between 77.3% and 82.6%. We would expect the resulting reduced feature sets to show patterns that permitted the MLAs to function this effectively, but again, there was no readily interpretable pattern. These findings are described in greater detail in Section 5.3.

Chapter 6 offers a discussion of these findings and explains them using CPR theory, which was introduced in Chapter 2. Chapter 7 takes up implications of this study, its limitations, including the question of statistical power, and suggestions for future study.

5.2 Findings from statistical analyses

Section 4.3.2 explained how data were collected and the 986 features were calculated. Section 4.4.1 described the methods by which the features were analyzed statistically. The full results appear in Appendix I, which reports for each feature the mean relative frequency with which Gender M and Gender F authors used the feature (including standard deviation); the "prevalent" gender, i.e., the gender representing authors who used the feature more frequently; and the Mann-Whitney U test p-value. This last is a

[2] Koppel et al. (2002) and Argamon, Koppel, Fine, and Shimoni (2003) make up what I have referred to as the Argamon/Koppel 02/03 study, some of the methods of which were models for the current study. See Section 3.3 for details.

calculation to determine whether the difference in mean relative frequency between the Gender M and Gender F authors is statistically significant. This study used an a priori α (or threshold p-value) of 0.05 as the boundary for statistical significance.[3] Thus only features where the p-values are less than 0.05 are deemed significantly different. For other features, Gender F and Gender M authors should be viewed as using the features with the same frequency. (They are in what pollsters sometimes refer to as a "statistical dead heat.") Nevertheless, the discussion below will occasionally comment on values where one gender or the other used a feature "non-signficantly" more frequently than the other gender.

See the discussion at Section 7.4 regarding my intention in the future to calculate effect sizes with regard to these values. Tests of statistical significance estimated the likelihood that the difference observed in the samples is "real" and not just the result of sampling. Effect sizes instead suggest whether the differences, if real, are of very great size (Utts & Heckard, 2006). As most of these features were not in Gaussian distributions, I have not performed the complex calculations to assess effect sizes for them. I intend to do so before publishing the findings of this empirical study.

The statistical findings are grouped into two subsections. Section 5.2.1 considers features that Argamon, Koppel, Fine, and Shimoni (2003) explored when concluding that male and female authors tended to write at different points on the "informational/involved" dimension identified by Biber (1995). Argamon, Koppel, Fine, and Shimoni (2003) found that men used more features associated with the "informational" end of the dimension, and women used more features associated with the "involved" end. The present study found that there were very few significant differences in the frequency of use of these features by Gender F and Gender M authors and that some of the significant differences contradict the earlier study.

Section 5.2.2 examines the statistical analysis of all the features (details of which appear in Appendix I) to determine whether Gender M and Gender F authors used them with significantly different frequencies. Table I.1 shows the subset of the full feature sets where there *were* statistically significant differences in frequency of use by Gender M and Gender F authors. Tables and text in Section 5.2.2 summarize some categories of features that generated some statistically significant differences. As this

[3] But see the discussion of the multiple comparison problem in the footnote at page 125.

chapter explains, there were some statistical differences that fell into certain categories, but many other features in the same categories were not significantly different based on author gender. These data do not support a finding of any pattern of differences between the Gender F and Gender M authors in these texts.

An example of each POS-bigram and -trigram discussed in this section, shown in at least one context in the papers collected for this study, appears in Appendix H.

5.2.1 Comparing and contrasting the Argamon et al. 2003 findings

The features Argamon, Koppel, Fine, and Shimoni (2003) identified as bearing a relationship to the informational/involved dimension described by Biber (1995) were examined for statistically significant differences between the Gender F and Gender M authors in this study. The values found here were compared with those found in Argamon, Koppel, Fine, and Shimoni (2003). The results are summarized in Table 5.1. As this section shows, hardly any of these features were significantly different in the present study, and where they were, they sometimes ran counter to the findings of Argamon and his colleagues.

Table 5.1: **Findings in Argamon et al. (2003) vs. present study**

	Argamon et al. 2003		**Present study**	
	Prevalence (non-fiction)	Mann-Whitney p-value	Prevalence	Mann-Whitney p-value
Nouns	(M)	*n.s.*		
NN (common sing.)			(F)	*n.s.*
NNS (common plural)			(M)	*n.s.*
NNP (proper sing.)			(F)	*n.s.*
NNPS (proper plural)			(F)	*n.s.*
Determiners	M	< 0.01	(M)	*n.s.*
Determiner+noun	M	< 0.01		
DT—NN (common sing.)			(M)	*n.s.*
DT—NNS (common plural)			(M)	*n.s.*
DT—NNP (proper sing.)			M	< 0.05

Prevalence indicates which gender used the feature more frequently. Mann–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for “not significant.” Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I; examples of bigrams in context appear in Appendix H.

Table 5.1: All function words... (continued)

	Argamon et al. 2003		**Present study**	
	Prevalence (non-fiction)	Mann-Whitney p-value	Prevalence	Mann-Whitney p-value
CD: Cardinal numbers	(M)	*n.s.*	(F)	*n.s.*
Attributive adjective	M	< 0.01		
JJ—NN			(M)	*n.s.*
JJ—NNS			(M)	*n.s.*
Prepositions	M	< 0.01		
IN: Preposition or subord. conj.			(M)	*n.s.*
Pronouns—all[4]	F	< 0.01	(M)	*n.s.*
First person	(F)	*n.s.*	(M)	*n.s.*
Singular	F	< 0.01	(F)	*n.s.*
Plural	(F)	*n.s.*	(M)	*n.s.*
Second person	F	< 0.01	(F)	*n.s.*
Third person	F	< 0.01	(F)	*n.s.*
Singular	(F)	*n.s.*	(F)	*n.s.*
Masculine	(M)	*n.s.*	(F)	*n.s.*
Feminine	F	< 0.01	(F)	*n.s.*
It	(F)	*n.s.*	(M)	*n.s.*
Its	M	< 0.01	(F)	*n.s.*
Plural	F	< 0.05	M	< 0.05
Verb, present tense	F	< 0.01		
VBP: Not 3rd p. sing.			(M)	*n.s.*
VBZ: 3rd p sing			M	< 0.01
Contractions	F	< 0.01	(F)	*n.s.*
Negation with "not"	F	< 0.05	(F)	*n.s.*

Prevalence indicates which gender used the feature more frequently. Mann–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I; examples of bigrams in context appear in Appendix H.

Table 5.1 presents the results of Argamon, Koppel, Fine, and Shimoni (2003) in the

[4] "Pronouns—all" for the current study refers to the frequency of the PRP tag (personal pronouns other than possessives).

left column. The findings of the present study appear in the right column. Recall that statistical significance in this study is evaluated using α=0.05. Thus, where $p < 0.05$ the difference between the two samples is deemed significant. Some researchers using this α use a second "level" or "degree" of significance, noting where $p < 0.01$. Though this does not make the difference in those instances "more significant," it does allow the researcher to assert with greater confidence that the difference is significant. Consequently, the results here show where the p-value is below 0.05 and where it is below 0.01.

The features are grouped into categories addressed by Argamon and his colleagues: Nouns, determiner-noun combinations, cardinal numbers, attributive adjectives, prepositions, pronouns, present-tense verbs, contractions, and analytical negation with "not." The significant differences in the present study supported the findings by the Argamon group in only one part of one of these categories: the determiner—proper noun bigram (DT—NNP). In two other categories—present-tense verbs and use of third-person plural pronouns—the significant differences in this study conflicted with those in the earlier study. In five of these categories—attributive adjectives (JJ—NN or JJ—NNS), determiners, prepositions, contractions, and analytical negation—there were no significant differences in the present study, but the non-significant differences corresponded by gender to the differences in the Argamon-group study. And in the remaining categories—nouns, cardinal numbers, and pronouns (other than those mentioned above), the present study's non-significant differences gave mixed signals, sometimes corresponding to the Argamon study and sometimes not. See Appendix H for examples of bigrams and trigrams in contexts where they were used by participants in this study.

Determiner+proper noun: Consistent with Argamon study

Argamon and colleagues found males used determiner-noun combinations more frequently ($p < 0.01$) than females did. This represented the only place in the current study where a significant difference went the same direction as in the Argamon study. But even here, the correspondence was not complete: Though Gender M participants in this study were significantly more likely to use the combination of determiner and proper noun (DT—NNP; $p < 0.05$), there were no significant differences in the frequency of use of determiners with common singular (DT—NN) or plural nouns (DT—NNS). Nevertheless, Gender M authors did use these features non-significantly more than Gender

F authors in the current study. See Appendix H for examples of bigrams and trigrams in contexts where they were used by participants in this study. [5]

Present-tense verbs and plural pronouns: Significant conflicts with Argamon study

The significant *conflicts* between the findings of Argamon, Koppel, Fine, and Shimoni (2003) and this study were in the use of the use of present-tense verbs and plural pronouns. Argamon and his colleagues found females used present-tense verbs more often than males ($p < 0.01$). The present study examined present-tense verbs, both third-person singular (which take the *-(e)s* ending for regular verbs; e.g., *talks* and *teaches*) and others, and found that Gender M authors used them all more than Gender F authors did, significantly so with regard to third-person singular forms ($p < 0.01$). As for third-person plural pronouns, Argamon and colleagues found that females used them significantly more frequently than males ($p < 0.05$). But in the present study, Gender M authors used them significantly more frequently than Gender F writers ($p < 0.05$).

Some categories corresponded without significance

Argamon, Koppel, Fine, and Shimoni (2003) found significant differences in the use of determiners (favored by males, $p < 0.01$), prepositions (favored by males, $p < 0.01$), analytical negation with "not" (favored by females, $p < 0.05$), attributive adjectives (favored by males, $p < 0.01$), and contractions (favored by females, $p < 0.01$). In each of these categories, the present study found the same gender preferences for the features, but none at significant levels. See Appendix H for examples of bigrams and trigrams in contexts where they were used by participants in this study.

[5] Dr. Serguei Pakhomov, a committee member reviewing this dissertation, suggested that the specific determiners used here might be of interest, particularly if they could be interpreted using the Givenness Hierarchy proposed by Gundel and others (Gundel, 2010). Givenness hierarchy theory attempts to explain the ways that Hearers associate referring expressions, like *these red books, the books*, or *they* to their referents in discourse, given the underdeterminacy of the referring expressions themselves. I'm interested in exploring this possibility further while preparing materials from this dissertation for publication.

Some categories conflicted, but also without significance

Finally, there are five categories where the present study found no significant differences, but its non-significant differences sometimes corresponded with the Argamon study and sometimes did not. These were nouns, cardinal numbers, and pronouns (other than third-person plural pronouns).

Recapping the comparison to Argamon et al. 2003

The very small number of statistically significant differences in the present study only once corresponded to the Argamon-group study and otherwise conflicted with it. The non-significant differences also sometimes ran the same way in the present study as in the Argamon-group study and sometimes not. Thus, as this discussion and Table 5.1 have shown, hardly any of the features used by Argamon, Koppel, Fine, and Shimoni (2003) to support the argument that men and women varied significantly along the informational/involved dimension of Biber (1995) showed significant gender difference in the same direction in the current study.

5.2.2 Signifcant differences in this study

Despite this study's failure to reproduce the statistically significant findings of Argamon, Koppel, Fine, and Shimoni (2003), and in particular the differences that Argamon and colleagues connected to the involved/informational dimension of Biber (1995), there were some features in this study where there were statistically significant differences between the Gender F and Gender M authors. These features are itemized in detail in Table I.1.

Many of the features where Gender M and Gender F authors varied significantly fell into four categories, which are further detailed in this section: Gender M authors used certain bigrams and trigrams with nouns more frequently than Gender F writers, and Gender M writers used certain quantifiers more. Gender F authors, on the other hand, used certain bigrams and trigrams including infinitive verbs more than Gender M, and Gender F authors used certain time and rhetorical sequencing adverbials more than Gender M writers. See Appendix H for examples of bigrams and trigrams in contexts where they were used by participants in this study.

Note, however, that it would be incorrect to say, for example, that Gender F authors used infinitive verbs more often than Gender M authors. Rather, of those places where the two gender's practices *did* vary, several happened to include infinitive verbs. This is an important distinction, which will be taken up further below.

Determiner-noun sequences in POS bigrams and trigrams

Among the features Gender M authors used significantly more frequently than Gender F writers were certain determiner-noun sequences. Table 5.2 shows the 71 features consisting of part-of-speech bigrams and trigrams that included a determiner followed by a noun. Of these, seven showed a statistically significant prevalence of use by Gender M authors (all at $p < 0.05$). In nearly half of these features was a determiner followed by proper noun (DT—NNP). So, for example, Gender M authors used determiner—proper noun significantly more frequently ($p < 0.05$).

But for a great majority of features including a determiner–noun sequence, there were no significant differences. There were 71 features in the corpus including a determiner followed immediately by a noun. Of these, however, 64 showed no significant difference between mean frequency of use by Gender M and Gender F authors, and for 25 of these features, Gender F authors used them non-significantly more frequently than Gender M authors.

Table 5.2: **Differences in features with determiner–noun components in present study**

Feature	Prevalence	Mann-Whitney p-value
Bigrams		
DT—NN	(M)	*n.s.*
DT—NNP	M	< 0.05
DT—NNS	(M)	*n.s.*
Trigrams		
CC—DT—NN	(M)	*n.s.*

Prevalence indicates which gender used the feature more frequently. Mann–Whitney *p*-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I; examples of bigrams and trigrams in context appear in Appendix H.

Table 5.2: Determiner–noun components... (continued)

Feature	Prevalence	Mann-Whitney p-value
CD—DT—NN	(F)	*n.s.*
Colon—DT—NN	(M)	*n.s.*
Comma—DT—NN	(F)	*n.s.*
Comma—DT—NNP	(M)	*n.s.*
Comma—DT—NNS	(F)	*n.s.*
DT—JJ—NNS	M	< 0.05
DT—NN—CC	(F)	*n.s.*
DT—NN—Colon	(F)	*n.s.*
DT—NN—Comma	(F)	*n.s.*
DT—NN—DT	(M)	*n.s.*
DT—NN—IN	(F)	*n.s.*
DT—NN—JJ	(F)	*n.s.*
DT—NN—MD	(F)	*n.s.*
DT—NN—NN	(M)	*n.s.*
DT—NN—NNS	(M)	*n.s.*
DT—NN—OQuote	(F)	*n.s.*
DT—NN—Period	(M)	*n.s.*
DT—NN—POS	(M)	*n.s.*
DT—NN—PRP	(F)	*n.s.*
DT—NN—RB	(M)	*n.s.*
DT—NN—TO	(F)	*n.s.*
DT—NN—VBD	(M)	*n.s.*
DT—NN—VBG	(M)	*n.s.*
DT—NN—VBN	(F)	*n.s.*
DT—NN—VBZ	M	< 0.05
DT—NN—WDT	(M)	*n.s.*
DT—NNP—CC	(F)	*n.s.*
DT—NNP—Comma	(M)	*n.s.*
DT—NNP—IN	(M)	*n.s.*
DT—NNP—MD	(M)	*n.s.*
DT—NNP—NN	(M)	*n.s.*

Prevalence indicates which gender used the feature more frequently. Mann–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I; examples of bigrams and trigrams in context appear in Appendix H.

Table 5.2: Determiner–noun components. . . (continued)

Feature	Prevalence	Mann-Whitney p-value
DT—NNP—NNP	(M)	*n.s.*
DT—NNP—Period	(M)	*n.s.*
DT—NNP—POS	(M)	*n.s.*
DT—NNP—VBD	M	< 0.05
DT—NNP—VBZ	(M)	*n.s.*
DT—NNS—CC	(M)	*n.s.*
DT—NNS—Comma	(F)	*n.s.*
DT—NNS—IN	(M)	*n.s.*
DT—NNS—Period	(F)	*n.s.*
DT—NNS—POS	(M)	*n.s.*
DT—NNS—RB	(F)	*n.s.*
DT—NNS—TO	(M)	*n.s.*
DT—NNS—VBD	(F)	*n.s.*
DT—NNS—VBP	(M)	*n.s.*
IN—DT—NN	(F)	*n.s.*
IN—DT—NNP	(M)	*n.s.*
IN—DT—NNS	(M)	*n.s.*
JJ—DT—NN	(M)	*n.s.*
NN—DT—NN	(M)	*n.s.*
NNS—DT—NN	(M)	*n.s.*
OQuote—DT—NN	(F)	*n.s.*
OQuote—DT—NNP	(M)	*n.s.*
RB—DT—NN	M	< 0.05
TO—DT—NN	(F)	*n.s.*
TO—DT—NNP	M	< 0.05
TO—DT—NNS	(M)	*n.s.*
VB—DT—NN	(F)	*n.s.*
VB—DT—NNP	(M)	*n.s.*
VB—DT—NNS	(F)	*n.s.*
VBD—DT—NN	(F)	*n.s.*
VBG—DT—NN	(M)	*n.s.*

Prevalence indicates which gender used the feature more frequently. Mann–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I; examples of bigrams and trigrams in context appear in Appendix H.

Table 5.2: Determiner–noun components... (continued)

Feature	Prevalence	Mann-Whitney p-value
VBN—DT—NN	(M)	*n.s.*
VBP—DT—NN	(M)	*n.s.*
VBZ—DT—NN	M	< 0.05
WDT—DT—NN	(F)	*n.s.*
WRB—DT—NN	(M)	*n.s.*
Total Determiner+Noun features: 71		
Total features with significant differences		**7**
Gender M prevalence	7	
Gender F prevalence	0	
Total without significant differences		**64**
Gender M prevalence	39	
Gender F prevalence	25	

Prevalence indicates which gender used the feature more frequently. Mann–Whitney *p*-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I; examples of bigrams and trigrams in context appear in Appendix H.

Sequences including infinitive verbs

The POS bigram sequence TO—VB represents the word "to" followed by the base or infinitive form of a verb (e.g., "to sue," "to be"). Among the features Gender F authors used significantly more frequently than Gender M authors were four representing infinitive-verb sequences: including the infinitive-verb bigram (TO—VB), and trigrams where common or proper noun or past-tense verb preceded an infinitive (NN—TO—VB, NNP—TO—VB, VBD—TO—VB). Table 5.3 displays these findings.

As with the Gender M preference for certain determiner-noun sequences, however, the infinitive-verb constructions Gender F authors used more frequently tell only part of the story. As Table 5.3 shows, there were 20 features in the corpus that included the

TO—VB bigram, but of these 16 showed no significant difference between the mean frequency with which Gender F and Gender M authors used the feature. Nine of these features in the corpus actually showed that Gender M authors used them non-significantly more frequently than Gender F authors.

Table 5.3: **Differences in features with infinitive–verb components in present study**

Feature	Prevalence	Mann-Whitney p-value
Bigrams		
TO—VB	F	< 0.05
Trigrams		
JJ—TO—VB	(F)	*n.s.*
NN—TO—VB	F	< 0.01
NNP—TO—VB	F	< 0.05
NNS—TO—VB	(M)	*n.s.*
RB—TO—VB	(M)	*n.s.*
TO—VB—DT	(F)	*n.s.*
TO—VB—IN	(F)	*n.s.*
TO—VB—JJ	(F)	*n.s.*
TO—VB—NN	(F)	*n.s.*
TO—VB—NNP	(F)	*n.s.*
TO—VB—NNS	(F)	*n.s.*
TO—VB—PRP$	(M)	*n.s.*
TO—VB—RB	(M)	*n.s.*
TO—VB—VBN	(M)	*n.s.*
TO—VB—Period	(M)	*n.s.*
VB—TO—VB	(M)	*n.s.*
VBD—TO—VB	F	< 0.01
VBN—TO—VB	(M)	*n.s.*
VBZ—TO—VB	(M)	*n.s.*
Total infinitive—verb features: 20		
Total features with significant differences		**4**

Prevalence indicates which gender used the feature more frequently. Mann–Whitney *p*-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I; examples of bigrams in context appear in Appendix H.

Table 5.3: Infinitive–verb components... (continued)

Feature	Prevalence	Mann-Whitney p-value
Gender F prevalence	4	
Gender M prevalence	0	
Total without significant differences		**16**
Gender F prevalence	7	
Gender M prevalence	9	

Prevalence indicates which gender used the feature more frequently. Mann–Whitney *p*-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I; examples of bigrams in context appear in Appendix H.

Use of function words classified as quantifiers

The use of function words that can be classified as *quantifiers* follows a similar pattern, with Gender M authors using some of them more frequently than Gender F authors did, but with many more features showing no significant difference or non-significant difference where Gender F authors used the feature more. Table 5.4 shows the findings. So, for example, Gender M authors used *all* ($p < 0.01$) and *many* ($p < 0.05$) significantly more frequently.

But most features that could be classified as quantifier function words showed no significant differences. Of 63 such features in the corpus, 59 exhibited no significant difference, and 27 showed non-significant differences where Gender F authors used the feature more; for example, Gender F law students used *eighteen*, *hundred*, *much*, and *six* non-significantly more frequently than Gender M authors did.

Table 5.4: **Differences in use of quantifiers**

Function word	Prevalence	Mann-Whitney p-value
again	(F)	*n.s.*
all	M	< 0.05
any	(F)	*n.s.*
both	(M)	*n.s.*
each	(M)	*n.s.*
eight	(M)	*n.s.*
eighteen	(F)	*n.s.*
eighteenth	(M)	*n.s.*
eighth	(F)	*n.s.*
eighty	(F)	*n.s.*
either	M	< 0.05
eleven	(F)	*n.s.*
eleventh	(F)	*n.s.*
every	(F)	*n.s.*
few	(F)	*n.s.*
fewer	(F)	*n.s.*
fifteen	(F)	*n.s.*
fifteenth	(M)	*n.s.*
fifth	(F)	*n.s.*
fifty	(M)	*n.s.*
first	(F)	*n.s.*
five	(M)	*n.s.*
forty	(M)	*n.s.*
four	(M)	*n.s.*
fourteen	(M)	*n.s.*
fourteenth	(M)	*n.s.*
fourth	(F)	*n.s.*
hundred	(F)	*n.s.*
less	(M)	*n.s.*
many	M	< 0.01
million	(M)	*n.s.*

Prevalence indicates which gender used the feature more frequently. Mann–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I.

Table 5.4: Quantifiers. . . (continued)

Function word	Prevalence	Mann-Whitney p-value
more	(M)	*n.s.*
most	(M)	*n.s.*
much	(F)	*n.s.*
neither	(M)	*n.s.*
nine	(F)	*n.s.*
nineteen	(M)	*n.s.*
ninth	(M)	*n.s.*
no	(F)	*n.s.*
none	(M)	*n.s.*
often	(F)	*n.s.*
once	(M)	*n.s.*
one	(M)	*n.s.*
quite	M	< 0.05
second	(M)	*n.s.*
seven	(M)	*n.s.*
seventh	(M)	*n.s.*
seventy	(M)	*n.s.*
six	(F)	*n.s.*
sixteen	(M)	*n.s.*
sixty	(M)	*n.s.*
some	(M)	*n.s.*
ten	(F)	*n.s.*
third	(F)	*n.s.*
thirteen	(F)	*n.s.*
thirty	(F)	*n.s.*
thousand	(M)	*n.s.*
three	(F)	*n.s.*
twelve	(F)	*n.s.*
twenty	(M)	*n.s.*
twice	(M)	*n.s.*
two	(M)	*n.s.*

Prevalence indicates which gender used the feature more frequently. Mann–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I.

Table 5.4: Quantifiers. . . (continued)

Function word	Prevalence	Mann-Whitney p-value
very	(F)	*n.s.*
Total quantifier features: 63		
Total features with significant differences		**4**
Gender M prevalence	4	
Gender F prevalence	0	
Total without significant differences		**59**
Gender M prevalence	32	
Gender F prevalence	27	

Prevalence indicates which gender used the feature more frequently. Mann–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I.

Time, sequencing, and rhetorical structuring function words

The last category of function words where the law students in this study showed significant differences in frequency varying with their genders is an inductive grouping that I created through a review of the significant differences in Table I.1. I have called this ad hoc grouping of function words *Time, sequencing, and rhetorical structuring words*. The shaded rows of Table 5.5 are the function words in this category showing significant differences. They include time and frequency adverbials, such as *today* and *tomorrow*, as well as *later*, *next*, and *soon*—and *often*, *generally*, and *rarely*. Also in this grouping are words used to sequence argument or discussion in the text, such as *firstly*, *finally*, and *thirdly*. Finally, this group included function words used to indicate the relation of the proposition in one clause to that in another clause or sentence, such as *consequently*, *furthermore*, *therefore*, and *moreover*. There were several such words that Gender F authors used more frequently. They included *furthermore* ($p < 0.01$), *afterward* ($p < 0.05$), and *secondly* ($p < 0.05$).

But the pattern previously observed with other categories of significant differences obtains again for this category: First, one function word—*then*—was used significantly more by Gender M authors in this study ($p < 0.05$). But Table 5.5 demonstrates that many function words in this category showed no significant difference in use frequency between Gender F and Gender M authors, and among non-significant differences, Gender M authors often showed a higher rate of use. There were 56 function words in this category present in the texts; of them only five showed a significantly greater use by Gender F authors, one showed a significantly greater use by Gender M authors, and 50 showed no significant differences at all. Even among the non-significant differences, 25 demonstrated greater frequency of use by Gender M authors and 25 by Gender F.

Table 5.5: **Differences in use of time, sequencing, and rhetorical structuring words**

Function word	Prevalence	Mann-Whitney p-value
accordingly	(F)	*n.s.*
after	(F)	*n.s.*
afterward	F	< 0.05
afterwards	(M)	*n.s.*
again	(F)	*n.s.*
ago	(M)	*n.s.*
already	(F)	*n.s.*
although	(F)	*n.s.*
always	(M)	*n.s.*
before	(F)	*n.s.*
besides	(M)	*n.s.*
but	(F)	*n.s.*
consequently	(F)	*n.s.*
despite	(M)	*n.s.*
during	(M)	*n.s.*
earlier	(M)	*n.s.*
early	(M)	*n.s.*

Prevalence indicates which gender used the feature more frequently. Mann–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I.

Table 5.5: Time, sequencing, rhetorical structuring. . . (continued)

Function word	Prevalence	Mann-Whitney p-value
eventually	(M)	*n.s.*
ever	(M)	*n.s.*
except	(F)	*n.s.*
finally	(M)	*n.s.*
firstly	(F)	*n.s.*
furthermore	F	< 0.01
generally	(M)	*n.s.*
hence	(F)	*n.s.*
however	(F)	*n.s.*
last	(F)	*n.s.*
lastly	(F)	*n.s.*
later	(M)	*n.s.*
moreover	(M)	*n.s.*
nevertheless	(F)	*n.s.*
next	(F)	*n.s.*
now	(F)	*n.s.*
often	(F)	*n.s.*
previous	(F)	*n.s.*
previously	(F)	*n.s.*
prior	(F)	*n.s.*
rarely	(M)	*n.s.*
secondly	F	< 0.05
since	(M)	*n.s.*
so	(M)	*n.s.*
soon	(M)	*n.s.*
still	(F)	*n.s.*
subsequently	(M)	*n.s.*
then	M	< 0.05
therefore	F	< 0.01
thirdly	(M)	*n.s.*
though	(M)	*n.s.*

Prevalence indicates which gender used the feature more frequently. Mann–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I.

Table 5.5: Time, sequencing, rhetorical structuring... (continued)

Function word	Prevalence	Mann-Whitney p-value
thus	(M)	*n.s.*
today	F	< 0.05
tomorrow	(F)	*n.s.*
twice	(M)	*n.s.*
until	(F)	*n.s.*
while	(M)	*n.s.*
yesterday	(F)	*n.s.*
yet	(M)	*n.s.*
Total time, sequencing, and rhetorical structuring features: 56		
Total features with significant differences		**6**
Gender F prevalence	5	
Gender M prevalence	1	
Total without significant differences		**50**
Gender F prevalence	25	
Gender M prevalence	25	

Prevalence indicates which gender used the feature more frequently. Mann–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I.

Recapping categories of statistical difference

Though many of the significant differences in mean frequency of use of certain features by Gender F or Gender M appear to fall into a small number of categories—determiner–noun sequences; infinitive verb sequences; quantifiers; and time, sequencing, and rhetorical structuring function words—these categories as a whole do not tend to show a preference of use by one gender over the other. For example, among 71 features in the corpus including a determiner followed immediately by a noun, 64 showed no significant difference between mean frequency of use by Gender M and Gender F authors, and for 25 of these features, Gender F authors used them non-significantly more frequently than Gender M authors. Table 5.6 displays a summary of the number of features in each of these four categories: determiner-noun sequences; infinitive verb sequences; quantifiers; and time, sequencing, and rhetorical structuring words. In all, there are 210 features in these categories, and only 21 of them showed significant differences between Gender F and Gender M authors, and in each category, the non-significant differences tended to run about evenly for Gender M and Gender F authors.

Table 5.6: **Summary of significant differences by feature category**

		Significantly different features			NOT significantly different			
Category	Detailed Table	Gender F Prevalence	Gender M Prevalence	Sub-Total	Gender F Prevalence	Gender M Prevalence	Sub-Total	Total Features
Determiner-noun components	Table 5.2	0	7	7	25	39	64	71
Infinitive verb components	Table 5.3	4	0	4	7	9	16	20
Quantifiers	Table 5.4	0	4	4	27	32	59	63
Time, sequencing, structuring	Table 5.5	5	1	6	25	25	50	56
Total		9	12	21	84	105	189	210

Numbers represent numbers of features in each category.

In short, this study does not support a claim that authors of either gender tended to use features in these categories more frequently than the other. It is interesting to note, however, that almost all the statistical differences in these categories tended to go toward one gender or the other. For example, all the significant preferences for the use of determiner–noun sequences was by Gender M authors. All the significantly different uses of infinitive verb constructions (those including TO—VB) showed a preference for Gender F authors. All the significant differences in the uses of quantifiers were by Gender M, and all the significant differences in the last category showed a preference for use by Gender F authors, except for the Gender M preference for the use of *then*. More research may shed greater light on the relationships of these differences to each other if any.

5.2.3 Findings regarding research questions 1 and 2

Among the research questions raised in Chapter 4 were the following:

- RQ1: Do Gender F and Gender M writers in a disciplinary genre in which they are being trained use lexical and quasi-syntactic stylistic features with relative frequencies that vary with their genders?

- RQ2: If so, do the differences appear in interpretable patterns?

The response to RQ1 appears to be "yes," there were certainly features in this corpus that were used significantly more by one gender or the other among the law students in this study. Table I.1 in Appendix I shows in detail which features were used significantly more frequently by one gender. There were more than 60 such features, including third-person singular present-tense verbs (VBZ, Gender M preference, $p < 0.01$), the infinitive verb bigram (TO—VB, Gender F preference, $p < 0.05$), and the trigram consisting of pronoun, third-person singular present-tense verb, and determiner (PRP—VBZ—DT, Gender M preference, $p < 0.05$).[6]

These differences did not not occur in an interpretable pattern, however, so we must respond "no" to RQ2. First, the "informational/involved" dimension used by Argamon, Koppel, Fine, and Shimoni (2003) (and originated by Biber (1995)) did not

[6] See Appendix H for examples of bigrams and trigrams described in this paragraph in contexts where participants in this study used them.

appear in this sample to vary with the gender of the law student authors. Second, though many of the features that exhibited significant differences varying with gender did fall into one or another of four categories—determiner–noun sequences; infinitive verb sequences; quantifiers; and time, sequencing, and rhetorical structuring function words—these categories as a whole did not tend to show a preference of use by one gender over the other.

Recall that the assessments of statistical significance in the differences in mean relative frequencies with which authors of different genders used the features in this study were made using an a priori α of 0.05.[7] Recall, too, that there were only 193 units (or *instances*, in machine-learning parlance) analyzed for each corpus here. To credibly maintain that there is no statistical difference between the frequency with which Gender M and Gender F authors used a feature, the researcher must establish that the sample size of the study is large enough to rule out (to a reasonable degree of probability) the possibility that lack of statistical difference between the two values is due to the size of the data sample; this is the principle of *statistical power*. If one observes a certain effect size—a difference in the mean frequencies—in a study with $n = 193$, the difference is less likely to be significant than if the same effect size were observed in a study with $n = 400$. Section 7.2 discusses further the concerns that a lack of power—that is, too small a sample size—may conceal differences that actually are significant within the population of the first-year law students at Academy School of Law and Lyceum Law College. Consequently, we can say that though the significant differences in this study did not appear in a meaningful pattern, further study would be necessary to demonstrate that there is no pattern of significant differences among the students in these law schools and among law students generally.

It would also be useful to calculate effect sizes for these differences. Unlike the test of statistical significance, which indicates the probability that the observed difference in the sample represents an actual difference in the population, the effect size is a calculation of how big the difference is. I have not calculated effect sizes for these features principally because of the difficulty of doing to with features that are not in a

[7] But see also the discussion of the multiple comparison problem in the footnote at page 125.

Gaussian distribution.[8] I have suggested in Section 7.2 that this may be an activity for the future.

In the meantime, however, Section 5.3 takes up the issue of whether machine-learning algorithms (MLAs) can classify these texts by author gender based on the features in this study. If they could, perhaps the features most useful to the MLAs could be shown to follow an interpretable pattern. Assuming that pattern included many features shown not to be statistically significantly different in this study, the MLA trials could identify features to focus on in future studies to offer sufficient statistical power to show a significant pattern of differences according to the epistemic tradition of statistics.

5.3 Findings from machine-learning analyses

Section 5.2 showed that there were statistically significant differences in the mean frequencies with which Gender F and Gender M authors used certain features, and even that many of those statistical differences fell into a few categories; nevertheless, they did not offer an interpretable pattern of variation because of the many features that were not used with significantly different frequencies. Chapter 4 noted that the features on which MLAs rely to classify texts are not necessarily different from each other in a statistically significant way, as they do not rely on the researcher's possibly arbitrary threshold of statistical significance to determine which features are useful for classification. It might thus be possible for the MLAs in this study to detect a pattern of differences more subtle than those recognized by tests of statistical significance. Chapter 4 also noted, however, that the models produced by MLAs may not be interpretable in the same way as statistical models, if they are interpretable at all. And in fact, this summarizes the findings of the MLA trials below.

Among the research questions raised in Chapter 4 were the following:

- RQ3: Can machine-learning algorithms categorize texts in a disciplinary genre—written by students being trained in it—by author gender based on the relative frequency with which they use lexical and quasi-syntactic stylistic features?[9]

[8] See the discussion in Section 4.4.1 regarding the lack of Gaussian distribution for most features in this study.

[9] Sections 3.4 and 4.2.3 provided an account of the rationale for classifying the authors in this study as "Gender F" and "Gender M" based on participants' gender self-identification.

- RQ4: If so, do they provide interpretable models?

Though three MLAs achieved accuracy here in classifying texts that was of both statistical and practical significance, the results do not yield a model for explaining the success of the MLAs.

Chapter 4 described the methods for collecting and preparing the data for this study in general, and Section 4.4.2 described the application of MLAs to the corpus in this study in an effort to determine whether the MLAs could classify the instances in this corpus by author gender using the features assessed for this study, 986 lexical and quasi-syntactic features. The corpus was subjected to classification by the following MLAs in the WEKA framework (Hall et al., 2009; "Weka 3: Data Mining Software in Java," n.d.; Witten et al., 2011), with each application of an MLA to the corpus described here as a "trial."

- `Winnow` (and balanced `Winnow`).
- Other linear models: `Logistic`, `SimpleLogistic`, and `VotedPerceptron`.
- Instance-based classifiers: `IB1` and `IBk`.
- Support-vector machines: `SMO` and `SPegasos`.
- Bayesian: `NaiveBayes` (with and without kernel estimator).[10]

In all, there were 32 trials. For each MLA, Appendix J includes a brief description of its method of functioning; the attributes with which it can function; necessary transformations of the data sets, if any, to implement the MLA; identification of parameters used with the MLA in the trials; and detailed results of the trials conducted with it.

Each MLA was applied to the instances in the corpus using all 986 features described in Section 4.3.2. The full feature set was then reduced using "scheme-specific" attribute selection, which identifies which attributes will maximize the success of that particular MLA. WEKA's `ClassifierSubsetEval` module with default configuration, including the `BestFirst` method for attribute selection, was used for attribute selection, except

[10] The kernel estimator is an alternative to basic `NaiveBays`, advised by Witten et al. (2011) for feature sets where continuous variables exhibit a non-parametric distribution. See the discussion in Section 4.4.1 regarding the lack of Gaussian distribution for most features in this study.

that Appendix J.5 describes a second round of attribute selection for the `NaiveBayes` classifiers. Thus, for each of the MLAs above, trials were run both with the full 986-feature data set and with a dataset created especially for that MLA by WEKA.

Recall, as Section 4.4.2 explained, that the performance of MLAs in this study was assessed by two criteria, one of *statistical significance* and one of *practical significance*. Statistical significance in these trials was assessed by comparing MLA performance to the *most frequent class baseline* (Jurafsky & Martin, 2009, p. 155) and assessing the significance of the difference using a paired t-test, with an a priori α of 0.05. This study begins with the a priori presumption that practical significance would be achieved only if the MLA classified texts with observed agreement of at least 66%.

This section summarizes the key findings of these trials.

5.3.1 Trials with full 986-feature data sets

In all, 15 trials were run with these MLAs using the full 986-feature data set. Details of all the trials appear in Appendix J. No MLA produced results that were significantly better than baseline with the full feature set.

5.3.2 Trials with reduced feature sets

There were 17 trials with MLAs using reduced feature sets. Details of all the trials appear in Appendix J. Of these trials, nine produced results that were significantly better than the baseline. These results are summarized in Table 5.7 on page 167. The four rows above the dividing line in Table 5.7 represent trials that achieved practical significance as defined for this study, which is to say, they had observed agreement exceeding 66%. Each row in the table represents one trial application of an MLA to the corpus.

There are two points worth noting regarding the outcomes of machine-learning classification shown in Table 5.7. These include performance of MLAs in the current study versus the findings in Koppel et al. (2002); and which MLA classifiers were most successful in this study.

Table 5.7: **Summary of all statistically significant MLA trials**

Machine learning algorithm	Observed Agreement
Practically significant trials	
NaiveBayes (WrapperSubset)	73.19%
SMO	71.57%
NaiveBayes w/kernel est. (WrapperSubset)	70.42%
SimpleLogistic	66.76%
Statistically (but not practically) significant trials	
Perceptron	65.29%
NaiveBayes (ClassifierSubset)	63.74%
IBk (instance-based)	62.93%
Spegasos	62.12%
Balanced Winnow (equal freq. 5 bins)	61.85%

Each row represents one trial. All trials showed statistically significant improvement over baseline (`ZeroR`, $p<0.05$). Trials below dividing line did not satisfy requirements for practical significance (> 66% observed agreement). All trials used reduced feature sets created with `ClassifierSubset` except those `NaiveBayes` data sets noted as having been made with `WrapperSubset`.

The machine-learning trials described in Koppel et al. (2002) achieved classification results with observed agreement that would be deemed practically significant in the present study. Koppel and his colleagues achieved observed agreement of 82.6% when classifying texts based on author gender using the balanced Winnow algorithm trained on non-fiction; and 79.5% observed agreement using it on fiction texts. (See Section 3.3.1 for details.) Only the WEKA implementation of `NaiveBayes`, trained on the corpus using the reduced feature set created by the `WrapperSubset` attribute selector in WEKA, achieved arguably similar results, with observed agreement of 73.19%.[11]

Of the four trials that achieved practically significant results, two used the WEKA implementation of a Naive Bayes classifier, `NaiveBayes`. Two other classifiers, `SMO` (a

[11] One trial with a separate corpus made up only of the fact sections of students' briefs achieved higher success: 76.94%. See Section 4.3.1 for a description of the components of the briefs. I have not reported analyses of portions of the texts that constitute separate corpora from the one studied here.

support vector machine) and `SimpleLogistic` (a linear-model builder), achieved one significant trial a piece.

5.3.3 The search for patterns in reduced feature sets

Recall that the reduced feature sets for machine-learning trials generally were created by the `ClassifierSubset` attribute selection algorithm in the WEKA machine learning framework. `ClassifierSubset` is a "scheme-specific" attribute selection algorithm, meaning that it produces a different selection of attributes or features for each algorithm designed to maximize that algorithm's performance. In theory, the result should be a subset of the 986-feature data set for each algorithm best suited to permit the algorithm to excel on the classification task.

Witten et al. (2011) also specifically recommend the `WrapperSubset` attribute selector for use with the `NaiveBayes` classifier. `WrapperSubset` is also a scheme-specific algorithm, but it uses a different approach for selecting optimal attributes. As a consequence of the recommendation of Witten et al. (2011), this study also included trials with `NaiveBayes` using attributes selected by `WrapperSubset`. These trials proved to have the best results. Both of these trials, one with the kernel estimator, and one without, produced practically significant results, making for two of the total of four practically significant trials.[12] What's more, `NaiveBayes` delivered the best classifier performance on the corpus using the feature set generated by `WrapperSubset`. Two other WEKA classifiers, `SMO` and `SimpleLogistic`, delivered practically significant results when classifying the corpus. They did so with feature sets generated for each algorithm by `ClassifierSubset`.

The question then is whether the successful MLA trials provide interpretable models of the differences between the writing of the Gender F and Gender M authors in this study. The answer appears to be 'no.'

Table 5.8 offers a detailed picture of the features that proved useful for these three algorithms. Each row represents one of the 70 features that appeared in at least one of the three reduced feature sets. The feature is identified in the left column; and feature rows are arranged by feature type: function words, parts of speech, POS bigrams, and POS trigrams. An example of each of the bigrams and trigrams in a context where

[12] See the footnote at page 165 for an explanation of the kernel estimator.

it was used by a participant here appears in Appendix H. Three columns identify the three successful MLAs—`SimpleLogistic`, `SMO`, and `NaiveBayes`. A bullet (•) appears in the column for an MLA for each feature that appeared in the reduced feature set for that MLA. For example, *again* and *forever* were among the features in the reduced feature set for `SimpleLogistic`, but neither of those words was in the reduced feature set for `NaiveBayes`. The final columns, on the right, permit comparison with the earlier statistical analyses by showing which gender used the feature more frequently and whether the difference between Gender F and Gender M's use of the feature was statistically different.

Table 5.8: **Features appearing in reduced feature sets of successful MLAs**

	Reduced feature sets			**Statistics**	
Feature	Simple Logistic	SMO	Naive Bayes	Gender prevalence	M-Whitney p-value
Function words					
'd		•		(F)	*n.s.*
again	•			(F)	*n.s.*
and			•	(F)	*n.s.*
because			•	(M)	*n.s.*
can			•	M	< 0.05
doubtful		•		(M)	*n.s.*
either			•	M	< 0.05
for		•	•	F	< 0.05
forever	•			(M)	*n.s.*
is			•	M	< 0.05
getting		•		(M)	*n.s.*
later			•	(M)	*n.s.*
many	•			M	< 0.01
my		•		(F)	*n.s.*
nothing		•		(M)	*n.s.*
ought		•		(M)	*n.s.*
possibly	•			(M)	*n.s.*
seventy		•		(M)	*n.s.*
somebody		•		(M)	*n.s.*
someone			•	(M)	*n.s.*
theirs	•			(M)	*n.s.*
then			•	M	< 0.05

Bullet (•) in a column indicates the feature was used in reduced features for the MLA associated with the column. Prevalence indicates which gender used the feature more frequently. M–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I. Examples of all bigrams and trigrams in use context appear in Appendix H.

Table 5.8: Features appearing in reduced feature sets... (continued)

	Reduced feature sets			Statistics	
Feature	Simple Logistic	SMO	Naive Bayes	Gender prevalence	M-Whitney p-value
under			•	F	< 0.05
undergo		•		(M)	*n.s.*
unlikely	•			(F)	*n.s.*
unusually		•		(F)	*n.s.*
where			•	(M)	*n.s.*
which	•			(M)	*n.s.*
who			•	(F)	*n.s.*
Part of speech					
Verb, past tense			•	(F)	*n.s.*
Verb, pres, 3rd p sing			•	M	< 0.01
POS Bigrams					
Comma—CC		•		(F)	*n.s.*
CD—CD		•		(F)	*n.s.*
NN—Comma			•	F	< 0.05
NN—TO		•	•	F	< 0.01
NN—VBZ	•		•	M	< 0.01
NNP—CC		•		(F)	*n.s.*
NNP—CD		•		(F)	*n.s.*
RB—IN			•	(F)	*n.s.*
VBZ—DT			•	M	< 0.05
POS Trigrams					
Comma—PRP—MD			•	M	< 0.05
CC—DT—NN			•	(M)	*n.s.*
CD—Comma—NNP			•	(M)	*n.s.*
DT—JJ—IN			•	M	< 0.05
DT—JJ—NNS			•	M	< 0.05
IN—NNP—NN	•		•	F	< 0.05
IN—PRP$—NNS			•	(M)	*n.s.*
NN—Comma—NNP			•	F	< 0.01
NN—IN—NNP			•	F	< 0.01
NN—NN—CC		•		(F)	*n.s.*
NN—NN—MD		•		(M)	*n.s.*
NN—TO—VB			•	F	< 0.01
NN—VBZ—IN			•	M	< 0.05
NNP—Comma—NNP		•		F	< 0.01
NNP—MD—VB		•		(M)	*n.s.*

Bullet (•) in a column indicates the feature was used in reduced features for the MLA associated with the column. Prevalence indicates which gender used the feature more frequently. M–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I. Examples of all bigrams and trigrams in use context appear in Appendix H.

Table 5.8: Features appearing in reduced feature sets… (continued)

	Reduced feature sets			Statistics	
Feature	Simple Logistic	SMO	Naive Bayes	Gender prevalence	M-Whitney p-value
NNP—NNP—Comma			•	F	< 0.05
NNP—NNP—IN			•	(M)	*n.s.*
NNP—TO—VB			•	F	< 0.05
NNP—VBD—TO			•	F	< 0.01
NNP—VBZ—IN			•	M	< 0.05
NNS—Comma—IN	•			(M)	*n.s.*
NNS—MD—VB			•	(M)	*n.s.*
POS—NN—TO			•	F	< 0.05
RB—IN—NN			•	(M)	*n.s.*
VB—IN—NNP			•	F	< 0.05
VB—VBN—TO			•	(M)	*n.s.*
VBD—TO—VB	•		•	F	< 0.01
VBZ—IN—DT		•		M	< 0.05
VBZ—RB—VBN			•	(M)	*n.s.*
WRB—DT—NN			•	(M)	*n.s.*
Summary of all features					
Best observed agreement of MLA	66.76%	71.57%	73.19%		
Number of features in reduced set	11	21	43		

Bullet (•) in a column indicates the feature was used in reduced features for the MLA associated with the column. Prevalence indicates which gender used the feature more frequently. M–Whitney p-value refers to result of Mann–Whitney U test: Significant at $p < 0.05$ or $p < 0.01$; or *n.s.* for "not significant." Where the difference between Gender F and Gender M is not significant, the prevalent gender is reported in parentheses: (F) or (M). Rows with findings *significant in this study* are shaded. Details of all means, standard deviations, and significance tests appear in Appendix I. Examples of all bigrams and trigrams in use context appear in Appendix H.

Here I will identify some observations about these data and attempt to interpret them in a way that might yield a knowledge model for the differences between Gender F and Gender M authors. Four observations relate to the differences between the feature sets that the three successful algorithms used. One relates to the features that the most successful algorithm—`NaiveBayes`—used.

First, the MLAs used different numbers and types of features to achieve their best performances. `SimpleLogistic` used 11 features to reach maximum observed agreement of 66.76%; `SMO` used 21 features to achieve 71.57% observed agreement; and `NaiveBayes` used 43 features to achieve 73.19% observed agreement. Seven of the 11 features that

`SimpleLogistic` used were function words. `NaiveBayes`, in contrast, used 43 features, of which 24 were part-of-speech trigrams.

Second, the MLAs relied to varying degrees on features that were significantly different in the statistical analysis. Recall that 65 of the 986 features studied in this project showed statistically significant differences in relative frequency of use between Gender F and Gender M authors. See Table I.1 for a complete list of the significantly different features, and the other tables in Appendix I for lists of all the features used in this study. Of those 65 significantly different features, 29 appeared in the reduced feature set for at least one of the three successful algorithms; but only five appeared in the reduced feature set for two of those algorithms: the function word *for* (Gender F prevalence, $p < 0.05$), the POS bigrams NN—TO (F, $p < 0.01$) and NN—VBZ (M, $p < 0.01$), and the POS trigrams IN—NNP—NN (F, $p < 0.05$) and VBD—TO—VB (F, $p < 0.01$). No feature appeared in the reduced feature set for all three algorithms.[13] Of the 29 significantly different features that appeared in at least one reduced feature set, `SimpleLogistic` used only four of them, `SMO` used only three, and `NaiveBayes` used 26 of them. It is perhaps not surprising that `NaiveBayes` used more statistically significant differences between the Gender F and Gender M writers to classify texts with the highest observed accuracy. But it tells us little about any patterns in these data that Section 5.2 has not already told us.

Third, the successful MLAs did not consistently use features in their reduced feature sets from categories like those discussed in the statistical analysis above. For example, `SimpleLogistic` used two function words classified as quantifiers in Section 5.2.2 (see Table 5.4 and accompanying text for details of the statistical analysis) in its reduced feature set: *again* (Gender F prevalence, *n.s.*) and *many* (M, $p < 0.01$). `SMO` used one quantifier: *seventy* (M, *n.s.*). And `NaiveBayes`, the champion for classification accuracy among the MLAs, used one quantifier: *either* (M, $p < 0.05$). Three of the 43 features that `NaiveBayes` used were infinitive—verb constructions (see Table 5.3 and accompanying text for details of the statistical analysis): NN—TO—VB (F, $p < 0.01$), NNP—TO—VB (F, $p < 0.05$), and VBD—TO—VB (F, $p < 0.01$). Only one

[13] Appendix H provides an example of each bigram or trigram in its use in context by a participant in this study.

of these, VBD—TO—VB, was used by one of the other two successful algorithms, `SimpleLogistic`.[14]

Fourth, the three successful algorithms do not appear to have uncovered any class or category of feature that is useful for categorization but that flew "under the radar" in the statistical analysis. As I noted earlier in this section, one virtue of MLAs is that they are not bound by the researcher's possibly arbitrary threshold of statistical significance when identifying features that are useful. The MLAs might be able to identify a consistently useful feature or group of features for distinguishing texts written by Gender F authors from those written by Gender M authors. The problem, on the other hand, is that the features that are useful for distinguishing the texts may defy any kind of theoretical categorization. Ideally, I would like to be able to say, "The Gender M authors in this study tended to do X," where X represents a class of features that fall into some broad category like those defined above (e.g., quantifiers, infinitive verbs, etc.). But in this case, I have not been able to identify any superordinate category(ies) in which to class the reduced features.

We are faced with the question of what the reduced feature sets *mean*, if anything; that is, what kind of knowledge representation they provide the researcher and whether it can be used to develop a theory about what is happening in these texts. Given that the reduced feature sets that proved useful for maximizing MLA performance do not exhibit even the limited degree of coherence seen in the statistical analyses above, the answer here may be that they do not provide a knowledge representation at all. These features may have proved useful to the MLAs because of circumstances arising from this particular data set—that is, they may be "overfitted" to the data, though cross-validation is intended to reduce the risk of overfitting.[15] Of course, exploring the inner workings of the attribute selection algorithms could shed more light on the results here, and further experimentation with attribute selection could perhaps result in better performance for the classifier algorithms and an interpretable model to support theory building. Section 7.2 takes up the question of whether the attribute selection process is a limitation of the present study and whether further research could resolve

[14] Appendix H provides an example of each bigram or trigram in its use in context by a participant in this study.

[15] See the discussion of overfitting in Section 4.4.2.

any related concerns. It also considers whether further study into the knowledge-making or epistemic utility of MLAs.

5.3.4 Findings regarding research questions 3 and 4

Among the research questions for the empirical study described in Chapter 4 were these:

- RQ3: Can machine-learning algorithms categorize texts in a disciplinary genre—written by students being trained in it—by author gender based on the relative frequency with which they use lexical and quasi-syntactic stylistic features?

- RQ4: If so, do they provide interpretable models?

We can now conclude that machine-learning algorithms can classify texts according to author gender, at least with reduced feature sets, at levels that have proven of practical interest in other studies. We continue to have little to go on though in claiming that there are meaningful patterns of difference between the Gender F and Gender M authors.

The machine-learning algorithms `SimpleLogistic`, `SMO`, and `NaiveBayes` classified the texts in this study with observed accuracy that was both statistically better than baseline and of practical interest, based on an arbitrary threshold that I established. Though it did not achieve the accuracy that Koppel et al. (2002) observed (between 79.5% and 82.6%), the `NaiveBayes` classifier achieved observed agreement of 73.19% using the reduced feature set created by the `WrapperSubset` feature selector. In any event, however, there was no clear pattern of differences between the Gender F and Gender M authors. The MLAs did not produce a theoretically interpretable model.

5.4 Conclusion

This chapter has presented from the findings of the empirical study described in Chapter 4. These two chapters addressed the following research questions:

1. Do Gender F and Gender M writers in a disciplinary genre in which they are being trained use lexical and quasi-syntactic stylistic features with relative frequencies that vary with their genders?

2. If so, do the differences appear in interpretable patterns?

3. Can machine-learning algorithms categorize the same texts by author gender based on the same features?

4. If so, do they provide interpretable models?

Section 5.2 offered the answers to the first two questions. First, it showed in Section 5.2.1 that the writers in this study did not exhibit the pattern of differences observed in Argamon, Koppel, Fine, and Shimoni (2003)—which placed men and women on the "informational"/"involved" dimension described by Biber (1995). Nevertheless, in Section 5.2.2, we learned that there were some significant differences between the Gender F and Gender M authors in this study. In fact, 65 features differed significantly in relative frequency between Gender M and Gender F authors. Of these features, 21 fell into four categories: determiner—noun components, infinitive—verb components, quantifier function words, and time, sequencing, and rhetorical structuring words. But 189 other features in those four categories showed no significant differences. Thus, those categories are not useful for distinguishing the writing of Gender F and Gender M authors in this study.

Section 5.3 offered the answers to the third and fourth questions. Again, there seemed some hope of success for characterizing differences between the writings of Gender M authors and Gender F authors, because three machine-learning algorithms were able to classify the texts correctly at levels that were both statistically and practically significant, achieving between 66.76% and 73.19% observed agreement. The MLAs were successful only when using reduced feature sets designed to maximize their performance. When I examined the features most useful for classifying these texts, however, I concluded once again that no patten of differences is evident.

Chapter 6 offers a discussion and explanation of these findings, and Chapter 7 takes up implications and limitations for this study, suggesting avenues for future research.

Chapter 6

Discussion: CPR theory and gender/genre

6.1 Introduction

This dissertation explores what I have called "cognitive pragmatic rhetorical" (CPR) theory and described in some detail in Chapter 2. In particular, it considers how CPR theory can contribute to our understanding of rhetorical and disciplinary, professional, and technical communication[1] theory and in particular to our theories of gender and genre performances. CPR theory takes account of a Writer's goals and beliefs about the world and explains the efforts the Writer employs in finding, discovering, or inventing her communicative performances with the principle of relevance, which holds that she will expend effort in her writing choices that is commensurate with the accessibility and strength of the goals for which she writes and the effect she expects her writing to have on the cognition of the Readers(s). In other words, relevance is a ratio of effect to effort. I've used this metaphor, thinking of relevance as a fraction with cognitive effects as the numerator—increasing it increases relevance—and cognitive effort as the denominator—increasing it decreases relevance—just as numerator and denominator effect the value of a fraction. Of course, relevance is assessed comparatively and not numerically.

[1] See page 2 for a definition and discussion of this term.

As a case study for applying CPR theory, the empirical study in this dissertation has examined the question of gender differences in the language of students completing their first year of training as law students. Sections 3.3 and 3.4.1 described a backdrop of studies using large (or larger) data sets generally showing there were differences in the writing of men and women (Argamon, Koppel, Fine, & Shimoni, 2003; Argamon, Koppel, Pennebaker, & Schler, 2007; Koppel et al., 2002; Rao et al., 2010; X. Yan & L. Yan, 2006). I commented at length in Chapter 3 on the methodological limitations of those studies—in particular their failure to make explicit their theories of gender and their methods for ascribing gender categories to participants and artifacts. Despite any methodological concerns, these studies resonate with a folk consciousness that is prepared to accept evidence of gender differences unquestioningly (DeFrancisco et al., 2014).

At the same time, a small number of earlier studies in professional and technical communication had shown no significant differences between the writing of female and male authors (Smeltzer & Werbel, 1986; Sterkel, 1988; Tebeaux, 1990). These studies had sample sizes that were too small—that is, they had too little statistical power—to infer from the lack of difference in them that there would be a similar lack of difference in a bigger sample or in the population at large; they also suffered the same sorts of methodological limitations as the more recent "big data" studies. Each of these studies examined student writing in professional or technical writing courses, and the authors concluded that the lack of gender differences there could be attributable to the students' efforts to "assimilate" to conventional professional language. Sterkel (1988) went as far as to suggest that the female authors had assimilated to a conventional professional style that evidenced a male communicative culture. I have frequently heard the same observation from researchers to whom I have presented portions of this research. I will take that issue up in some more detail in Section 7.4.

These two groups of studies—big-data studies that showed gender differences in texts without a common purpose or genre, and smaller studies that showed no gender differences in texts written with a common purpose—prompted me to ask whether gender and genre interact, and if so, how. I wanted to see if gender differences of the kind previously identified would appear in a larger sample of texts that were controlled for text genre and attempt to offer a theoretical explanation for those findings, whatever they might be. That led to the empirical research study in this dissertation, the methods for which were described in Chapter 4 and the findings in Chapter 5. That study asked and provided answers to four research questions:

1. Do Gender F and Gender M writers in a disciplinary genre in which they are being trained use lexical and quasi-syntactic stylistic features with relative frequencies that vary with their genders? The answer is "yes."

2. If so, do the differences appear in interpretable patterns? The answer is "no."

3. Can machine-learning algorithms categorize the same texts by author gender based on the same features? The answer is "yes."

4. If so, do they provide interpretable models? The answer is "no."

Recall that Sections 3.4 and 4.2.3 provided an account of the rationale for classifying the authors in this study as Gender F and Gender M based on participants' gender self-identification. As Chapter 5 showed, there were some significantly different features between the writings of Gender F and Gender M authors, and machine-learning algorithms (MLA) were able to classify the texts with observed agreement between 66.76% and 73.19%—levels comparable to some of the previous studies. But significantly different features and the features most useful for MLA classification did not fall into readily identifiable patterns as previous studies suggested they would. So, for example, Gender F authors used four features including the part-of-speech sequence TO—VB, representing the word "to" followed by the base or infinitive form of a verb (e.g., "to sue," "to be"), significantly more frequently than Gender M authors did in the corpus I studied. Nevertheless, there were 16 other features including this sequence where there were no significant differences in relative frequency between the Gender F and Gender M authors. And as for many of those features, Gender M authors used them non-significantly more frequently than Gender F authors. It is thus impossible to claim based on this evidence that Gender F authors used infinitive verb constructions more frequently than Gender M authors. (See Section 5.2.2 for a fuller presentation of these findings.)

Previous studies, and particularly Argamon, Koppel, Fine, and Shimoni (2003) and Koppel et al. (2002), suggested that patterns of difference between male and female authors exhibited characteristics of the "informational"/"involved" dimension described by Biber (1995). On this account, women are associated with involved writing, which displays "interaction between the speaker/writer and the listener/reader, such as first and second person pronouns"; other "prominent characteristics" of involved writing are "analytic negation [negation with 'not'], contractions, and present-tense verbs" (Argamon, Koppel, Fine, & Shimoni, 2003, p. 332). And on this account, men are more likely to write using features of the informational end of the dimension, exhibiting larger numbers of specifiers and particular types of prepositional phrases.

This chapter will take the patterns of gender difference described in the previous, big data studies as a given—despite methodological questions—in part because of the popular belief that men and women do communicate differently. The work of this chapter then is to explain why the writers who participated in the study in this dissertation did not exhibit these (or any other discernible) patterns of difference. CPR theory provides a starting point for this discussion. Section 6.2 provides a brief reprise/overview of CPR theory, discussing the CPR-theoretic production and interpretation procedures in more detail than Chapter 2 did. Section 6.3 discusses the social context for the current study, noting the data that would be useful and necessary to perform a complete CPR-theoretic analysis. Section 6.4 provides a CPR-theoretic account of gender, and Section 6.5 provides such an account of genre knowledge. Finally, Section 6.6 synthesizes these accounts with the findings in the empirical study in this dissertation.

6.2 Reprise of CPR theory

Chapter 2 describes cognitive pragmatic rhetorical (CPR) theory in some detail, including its evolution from theories of rhetoric, experimental pragmatics, and cognitive science. This section provides an overview of CPR principles and the data a CPR theorist would examine when performing a full CPR-theoretic analysis. Because data collection for the empirical portion of this study did *not* include gathering a wide variety of data that would be useful for using the CPR model, I will note many tentative factual assertions—hypotheses in a way—based on my anecdotal experiences and intuitions as a teacher for eight years of the sort of classes in which this study took place. I will take up the limitations suggested by the hypothetical nature of these assertions, and the directions for future study that they suggest, in Chapter 7. In this section, I will extend the discussion in Section 2.5.2 to explore the CPR-theoretic production and interpretation procedures and some examples of their application. In Section 6.3, I'll place information from this study in the context of CPR theory.

At the heart of CPR theory is the concept of relevance. Relevance is the ratio of cognitive effects to the cognitive effort expected to be necessary to produce them. The greater the effect, the greater the relevance; the greater the effort, the lesser the relevance. I adopt the claim of Wilson and Sperber, 2006, p. 610, that "[h]uman cognition tends to be geared to the maximization of relevance." Effects and effort are measured in the context of the human agent's goals and assumptions while she is engaged in producing or interpreting an utterance (or text), that is, in the context of her cognitive environment. For an individual agent at a given time, her cognitive environment is the union of the set of assumptions and goals accessible to the agent and her emotional state.[2] Assumptions are thoughts that can be expressed in propositional form and are treated by the individual agent as representations of the world, including the states of mind of the agent or others (meta-representations) and including hypothetical propositions. Goals are consequences (end states or otherwise) desired or unwanted by an agent capable of motivating an agent to action.

But an agent's cognitive environment also includes the imputed cognitive environment of her audience, consisting of the assumptions that the agent has about the cognitive environment of another agent or group of agents, including the agent's assumptions about the cognitive environment that is being imputed to her. In other words, these are the assumptions that the agent believes she and other the other agents all share about their current situation.

The Speaker or Writer no doubt has many goals that motivate her behavior and a large store of beliefs—assumptions—about the world, herself, and her audience. In the process of communication, some goals and assumptions are likely more accessible to her than others, and she is likely more committed to some than others. More accessible goals and assumptions are available to influence information processing and decision-making with lower search costs than other assumptions or goals. In general, search costs are lower where a cognitive heuristic or the agent's habitual practice leads to the

[2] Note I have generally not taken up the issue of emotional states in this dissertation. While I believe a discussion if it is essential for CPR theory to be a general theory of text production and interpretation, I will leave that matter to Section 7.4 and future studies.

assumption or goal, or the assumption or goal is closely related to one already made accessible by previous communication. Commitment measures the extent to which the agent is committed to the goal or assumption; I have sometimes referred to the "strength" of a goal or assumption in this context. Accessibility and commitment can interact in a variety of circumstances. For example, though I may be firmly committed to losing 20 pounds in the coming two years, in a conference with a faculty colleague about a grading dispute with a student, I will be prepared to expend considerable effort to resolve the dispute. Overall, I may be less committed to resolving the grading dispute than losing the weight, but the conference with the colleague makes my goals about the grading dispute more accessible.

A Speaker seeks by her utterance to change the cognitive environment of the Hearer. Perhaps she seeks to inform him by adding to his store of assumptions, getting him to abandon an assumption he previously adopted, or strengthening or weakening existing assumptions. Perhaps she seeks to move him by getting him to adopt a new goal, to abandon an existing goal, to make an existing goal more accessible, or to increase or decrease his commitment to an existing goal. Finally, perhaps she seeks to delight him or operate on his emotions in some other fashion, by making him feel happy, sad, insecure, loved, or friendly toward the Speaker. Which of these changes the Speaker seeks to work on the Hearer's cognitive environment will depend on which of the Speaker's goals are most accessible to her when she is speaking and the strength of her commitment to them. The effect the Speaker hopes to achieve in a communicative exchange is thus a function of the accessibility and strength of her goals and her expectation that she can advance those goals by changing the cognitive environment of her Hearer.

If the change she wishes to work in the cognitive environment of the Hearer is essential to achieving a goal that is very accessible and very important to the Speaker, it has high expected effects. If on the other hand she does not expect to gain much from an interaction with the Hearer, it will have low expected effects. Consider the following contrasting situations: Imagine that the Speaker and her sick child are hostages, and the Speaker is attempting to persuade the Hearer—her captor—to release the child for treatment. Contrast the situation where the Speaker passes a colleague in the hall, and the colleague has said "Good morning." In the former case, the Speaker's goal is clear, accessible, and one to which she is strongly committed; and she must hope that her communicative performance can advance this goal. Her communicative performance here has high anticipated effects. In the latter, case, the Speaker may have a generalized goal of maintaining or improving her affiliation with the colleague, but the hallway interaction is unlikely to play a major role in achieving that goal. The interaction has low anticipated effects.

The effort side of the relevance ratio is the search cost associated with extending her own cognitive environment to make assumptions and communicative strategies available to her. Given what (she thinks) she knows about the Hearer's cognitive environment, she might need to "dig deep" to make accessible to herself communicative options: "What will appeal to him in this situation to get him to respond as I wish?" CPR theory thus measures production effort in terms of search costs: The greater the extent to which she can use habitual utterance patterns or utterance patterns grounded in automatic cognitive heuristics, the lower the effort required to generate her utterance. The more she has to search and find options for constructing her utterance—perhaps by closely considering the cognitive environment she imputes to the Hearer, perhaps by spending time considering ways that she could use rational and emotional appeals to make assumptions or goals accessible to the Hearer that will advance the Speaker's goals—the higher the effort required. Returning to the previous examples, the Speaker addressing the hostage-taker might carefully consider assumptions she has about him. For example, if he is a co-worker, angry that he has lost custody of his own children, she may conclude that he may respond to arguments appealing to his emotions about his children. At the same time, she may conclude that the exact choice of words she uses may be critical to ensuring she arouses his sympathy rather than intensifying his anger. The colleague in the hall, on the other hand, may not give a second's thought and respond "Good morning!"

The Speaker's performance should also attract the Hearer by satisfying *his* expectations of relevance according to the communicative principle of relevance, which appeared on page 31. With these components in place, I can offer the CPR-theoretic production procedure:

(1) *CPR-theoretic production procedure:* The Speaker should incur search costs in producing her utterance proportional to the effect she expects it to have on

> the Hearer's cognitive environment, taking into account the imputed cognitive environment, the Hearer's likely assessment of the utterance's relevance to the Hearer, and the accessibility and strength of the goal the Speaker is attempting to advance.

Note that this procedure presumes that the Speaker can assess at the outset what level of search costs she should invest in producing the utterance; she then produces the utterance.

Relevance for the Hearer consists of the ratio of effects to efforts, too. But unlike the Relevance Theory of Sperber and Wilson (1995) ("SWRT"), CPR theory does not hold that a Hearer's assessment of cognitive effects is limited to the enhancement of the assumptions available to the Hearer. On the contrary, CPR theory acknowledges that the Hearer may be engaged in the communicative exchange to obtain something more or other than information. He may seek an interaction that entertains him; that evokes emotional responses such as fear, affection, anger, or camaraderie; or that provides him motivation to take some step to advance his goal. Consider the following examples: The Hearer reads a news article to learn about the personal history of a political candidate. He watches the movie *Animal Crackers* in order to experience amusement. He attends a vigil commemorating the accidental death of local teens in hopes that he can release feelings about, or make sense of, the event. He talks to his personal trainer in order to be motivated to work out more frequently or regularly. As with the Speaker, the significance of the effects obtained from a Hearer interpreting an utterance depend upon his goals in taking part in the interaction in the first place. The expected effects of interpreting utterances are proportional to the accessibility of the Hearer's goals for engaging in the communicative exchange and his commitment to them.

CPR theory also measures the Hearer's effort in terms of search costs: The greater the extent to which he can identify the Speaker's utterance as an example of a habitual utterance pattern, the lower the effort required to interpret her utterance. The greater the extent to which the Hearer can rely on assumptions that are in the cognitive environment he imputes to the Speaker or in the Hearer's own cognitive environment, the lower the effort of interpretation. The more the Hearer has to search and find less accessible assumptions and goals as resources for interpreting the utterance, the higher the effort required. As when the Speaker produces an utterance, the Hearer may need to engage in search effort by considering the cognitive environment he imputes to the Speaker, perhaps by spending time considering assumptions on which she may be relying, or goals that she may be attempting to advance that he has not imputed to her before. He may need to consider her emotional state or his own.

However, because the Hearer can presume that the Speaker has attempted to make her ostensive communication optimally relevant to him, he can continue generally to rely on the SWRT comprehension procedure, which directs him to follow a path of least effort in interpreting the utterance and to stop when his expectations of relevance are met. The CPR theoretic comprehension procedure accounts for the accessibility and commitment of the Hearer's goals, though, and can thus be defined this way:

(2) *CPR theoretic comprehension procedure:* The Hearer should follow the path of least effort in interpreting the Speaker's utterance, testing interpretative hypotheses in order of accessibility, stopping when his expectations of relevance are satisfied, taking into account the imputed cognitive environment and the goals the Hearer is attempting to advance.

Let's consider some examples, beginning with one where the Hearer has a very strong goal, oriented toward gathering information. Imagine the complement to the scenario described above: The Hearer and his sick child are being held hostage in the workplace by an angry co-worker, disconsolate over having been deprived of custody of her own children. In this situation, the Hearer may invest considerable effort in applying rational inference to any utterance of the Speaker, in hopes that he might glean some bit of information that will be helpful in persuading her to release his child for treatment. Here, his goal of protecting his child is highly accessible and strong.

Consider a lighter-hearted pair of examples: Imagine you are on a fancy safari in Namibia, and the guests and tour guide are sitting at breakfast in the fancy pajamas provided to you by the tour company. The tour-guide is giving you a safety talk, explaining how dangerous some of the large animals are and that it is sometimes necessary to use deadly force against them in self-defense. She mentions that the dangers are greatest in the early-morning hours. She then utters (3).

(3) One morning I shot an elephant in my pajamas.

Now, consider instead that you are a fan of Groucho Marx and are attending a showing of the movie *Animal Crackers*, which stars the Marx Brothers but which you have somehow never managed to see. In one scene, Groucho utters (3).

The sentence in (3) represents an instance of syntactic ambiguity called "attachment ambiguity" (Jurafsky & Martin, 2009, p. 432). In one variety of attachment ambiguity, a prepositional phrase such as "in my pajamas" could be said to modify or attach to two different locations in the sentence. Here, it is possible that "in my pajamas" is meant to attach to the pronoun "I," meaning that the Speaker was in her pajamas when she shot an elephant. It is also possible that the "in my pajamas" is meant to attach to the object noun phrase in the sentence: "an elephant in my pajamas," that is, meaning the elephant was wearing the Speaker's pajamas. According to the principles of pragmatics, you must use inference to determine where the prepositional phrase attaches so you can interpret the utterance.

In the first instance, you listened to your tour guide, who seems to know her stuff, describe what sounded like actual events that were important for your safety, while you all sat in your pajamas. We will assume, for the moment, that the hypothetical assumption, 'Elephants sometimes wear pajamas,' is not accessible in anyone's cognitive environment. When she utters (3), the most accessible interpretation of the role of "in my pajamas" in the sentence supports the assumption expressed by (4):

(4) 'One morning, the tour-guide—while still wearing her pajamas—shot an elephant.'

This interpretation is most consistent with your assumptions and goals at this time.

In the second instance, you are a fan of Groucho Marx, and you know that word play is at the heart of his comedic appeal. Whenever Groucho says anything, you anticipate that it may be a play on words or a joke. It is likely that *as he is saying his line* in (3), you are already considering alternatives to the typical interpretation of the utterance. The accessibility of, and your commitment to, your goal of being entertained are a measure of your anticipated effect, which licenses greater effort for interpreting the utterance. Likely, you will not just be interpreting (3) while it's being uttered, you may be anticipating (5), perhaps supplying the punchline yourself. This may be why Groucho allows less than a second to expire between uttering (3) and uttering (5)

(5) How he got in my pajamas, I don't know.

6.3 CPR theory in context in this study

For the researcher, setting the stage for a CPR theoretic analysis of a Speaker's production of text means gathering a variety of facts about the Speaker's cognitive environment. Many of these data I did not gather from participants in this study, but I will supply suspicions and hypotheses about them here based upon my experiences and intuitions from having taught courses of the kind in which the study texts were produced; this includes intuitions I have about the instructors based upon my experience practicing in the legal field and counseling students as part of a structured writing group at one of the law schools in the study (though I did not teach or counsel at either of the schools during the year of this study). All these assertions should be subjected to empirical scrutiny, but those tasks are beyond the scope of this dissertation, though I propose such extensions of this study in Section 7.4.

The cognitive environments of the participants in this study included their accessible goals. These legal writing courses were framed as an important part of students' first-year educations. I suspect attorneys, law students, and (most) law professors recognize that training in presenting legal reasoning and argumentation in written form is essential for an attorney to succeed. All the following are probably true: (1) Students' performance on the year-end brief assignment played a significant role in their legal writing grade for the semester or the whole year. (2) Students were counseled that their year-end briefs would likely be their writing samples for job and internship applications, which begin usually in the fall of their second year. (3) As students' legal writing professors were likely the only instructors during their first year to have worked with the students closely, students probably expected to seek letters of recommendation from these instructors; students likely recognized that performance on this assignment would affect the type of letter of recommendation the legal writing instructor would be willing to provide the student. All these factors—if true—speak to students' commitment to, and the accessibility of, the students' goals of doing well on this assignment.

Note, however, that legal writing was only one of many courses these students took in law school in their first year. Students had to balance the demands of all courses, as their grades and grade-point averages may also have played an important role in employers' decisions to hire. The demands of other courses may thus decrease the accessibility of the goal of success in legal writing. Students' individual situations, including relationships and home life, may also provide competing goals that obtain higher levels of commitment from the students or are more accessible to them, at least from time to time. About these it is more difficult to speculate or hypothesize.

I suggested in the text beginning at page 81 that signaling the Speaker's sex through gendered communicative performances is commonly an accessible goal of the Speaker and that assumptions about a Speaker's sex usually have a high effect-value to the Hearer. I speculate though that within the communicative situation of these assignments, the students' goal of communicating gender identities consistent their sex is relatively *inaccessible* compared to the goals of performing well on the writing task. This assertion should be subjected to empirical scrutiny.

Assumptions students hold about the writing task play an important part in their cognitive environments. Normally, among the assumptions that the CPR theorist would examine are a Speaker or Writer's accessible typifications; in other words, whether she has at her disposal assumptions about what performances are fitting or appropriate in certain communicative interactions. Keep in mind that at this stage, most students had probably seen very few texts in the genre they are being asked to write. They may have had some textbook examples or examples supplied by the instructors. But in my experience it is not the instructional model of these legal writing programs to analyze or discuss the generic characteristics of legal memoranda of this kind before asking students to write them. The students had nevertheless spent more than six or seven months reading court opinions and other legal texts that signal a wide variety of lexical and quasi-syntactic preferences in legal writing generally. Students had probably submitted earlier writing assignments during the year and received feedback from instructors about their writing styles, in the form of grades, specific comments, or both. In my experience, students begin the first year of legal writing by attempting to mimic the tendency of some judges to write "legalese." Students readily identify these variations from common stylistic usage and attempt to employ them themselves, sometimes with comic consequences. Over the course of the year, legal writing instructors discourage legalese as well as stylistic habits—like use of the passive voice—that students may have brought with them from their previous careers.

But students' assumptions can also be uncertain at this stage. For example, if they had seen few examples of a certain type of legal writing, they may not have had confidence about what formal elements of it are typified. This uncertainty also often surrounds the students' efforts to construct the imputed cognitive environment of their instructors. For example, students might receive conflicting advice from a textbook and an instructor on a stylistic matter. They might even receive conflicting advice from different instructors. Consider this example that I witnessed while tutoring students: some instructors would strongly prefer two sentences like those in (6) over those in (7):

(6) Plaintiff Chris Smith ("Mr. Smith") brings this action against Gourmet Design, Inc. ("Gourmet Design"), to enforce a contract between them. Mr. Smith and Gourmet Design executed an agreement under which Gourmet Design would develop a commercial barbecue sauce recipe for Mr. Smith in return for a payment of $45,000.

(7) Plaintiff Chris Smith brings this action against Gourmet Design, Inc., to enforce a contract between them. Mr. Smith and Gourmet Design executed an agreement under which Gourmet Design would develop a commercial barbecue sauce recipe for Mr. Smith in return for a payment of $45,000.

Other instructors would strongly prefer (7). Assuming that a student has Professor A, and the student knows that Professor A prefers (7), the student has an assumption about how to handle this type of situation. Nevertheless, the student may read a textbook example or other examples of legal writing that show a preference for (6). My own unsystematic review of actual, filed court memoranda provides evidence for both of these approaches being used by otherwise apparently competent lawyers. Thus, the student may have accessible to her an assumption like (8), but she may be only lightly committed to it owing to the conflicting evidence at her disposal.

(8) 'Legal writers prefer that abbreviated names of parties in legal briefs should be handled as in the example in (7).'

A more sophisticated student might have accessible to her assumptions like (9) and (10).

(9) 'Professor A prefers that abbreviated names of parties in legal briefs should be handled as in the example in (7).'

(10) 'Some experienced legal writers (including perhaps the textbook author or some practicing attorneys) prefer that abbreviated names of parties in legal briefs should be handled as in the example in (6).'

This student might be able to adopt these assumptions with a great degree of commitment, based on the evidence she has. Of course, given the great number of stylistic topics about which she may be forming such assumptions during law school, there is always the risk that she will not have the right assumption accessible to her when she is composing: "I can't remember, does Professor A prefer (6) or (7)?"

The student is of course—to a greater or lesser degree, depending on her commitment to the goal of legal-writing success—searching her assumptions and inferences she can draw from them to construct other elements of the instructor's cognitive environment. What are his goals, his assumptions about this task? What is his emotional state? I don't mean to suggest that the entire educational experience of law students involves the armchair psychology of trying to read instructors' minds, except to the extent that CPR theory holds that all human communicative activities require such mind-reading.

In addition to the Writer's cognitive environment in the form of goals, assumptions (and typifications), and the imputed cognitive environment that she infers about the Readers, the CPR researcher should also consider unconscious heuristic resources, including the Writer's cognitive biases and habitual communicative practices. The Writer or Speaker may not be aware of these in any conscious way. This study did not attempt to assess the Writer's habitual stylistic habits in their writing at the beginning of their law-school training, though Section 7.4 suggests that as an avenue for future research. As for the students' habitual communicative practices that may vary with their genders, that is the topic of the next section.

6.4 CPR theory accounts for gendered language

In Section 3.4.2, I briefly made assertions about gender performances that I claimed were consistent with CPR theory. I first defined gender this way:

(11) Gender is a loosely and culturally defined set of social behaviors that are expected to make it possible to distinguish the two most common sexes from each other.

I speculated that a cognitive bias in humans makes it a goal of high accessibility and commitment for an agent to ascribe a sex category to other agents as soon as possible in their interactions. As I have explained earlier, this is likely an important adaptive cognitive bias in species that reproduce sexually. Even among members of the species who are not heterosexual or who are past the age of fertility, it is likely that the cognitive bias applies. From a CPR-theoretic perspective, then, assumptions about a Speaker's sex have a high effect-value to the Hearer. A complementary speculation is that each agent has a strong, accessible a goal of signaling her (or his) sex by means of gender performance. The gender performances of an agent allow others to infer assumptions about the agent's sex. Generally, this involves very low effort on the part of the agent: Because rightly or wrongly (mostly wrongly) certain social roles are allocated to and associated with persons based on their gender performances, and these roles begin to be allocated at early ages, agents' gender performances are likely some of the most deeply and habitually ingrained in their communicative repertoires. They come with low search costs; that is, they require low effort.

High effect and low effort make for high relevance, both for Speaker and Hearer, when it comes to gender performances. Divergences from expected gender performances are jarring, because they require higher effort and no apparent gain. Conventional gender-performance practices are likely very habitual and unconscious, though some have claimed that they can also be strategically deployed; see for example, DeFrancisco et al. (2014).

Whether it is makes sense to exhibit or suppress habitual gender performances will be a relevance-based assessment. In general, where there is no contextual reason for the agent to vary from habitual performances, we would expect habitual performances.

This supports the theory of Butler (1997), who argued that gendered performances produce and reproduce gender. Repetition of gender performances reinforces their habitual nature, which reinforces their repetition. There are circumstances, however, where the likely effects of gendered performances on the audience are less certain and other assumptions and goals in the Speaker's cognitive environment will guide her stylistic choices. Varying from habitual practices requires considerable cognitive effort, probably conscious attention. Conforming to a professional genre is likely one of these circumstances.

6.5 CPR theory accounts for genre knowledge

In Section 3.5.6, I made some brief observations about genre and its conception under CPR theory. I offered a definition of genre:

(12) A genre is a loosely and culturally defined set of communicative behaviors, usually formal conventions, a Speaker or Writer expects to have a particular effect or effects on a Hearer or Reader, based on assumptions about a typified situation in the Speaker's imputed cognitive environment.

Each time an agent encounters a communicative pratice, she must decide whether to "typify" the text and the social situation in which it operates and determine which formal characteristics of the text are essential to its social function. Cognitive pragmatic rhetorical (CPR) theory predicts that these interactions will be governed by the principles of relevance, and that the agent will use the CPR-theoretic production and comprehension procedures.

First, under CPR theory an agent's genre knowledge is just a "slice" of her cognitive environment. For any given agent at a given time, her cognitive environment consists of her assumptions (representations of factual knowledge about the world and the minds of herself and those around her) and goals, weighted depending on how accessible they are to her and how committed she is to them, and her emotional state. Particularly important for genre knowledge is that subset of her assumptions relating to the imputed cognitive environment, the assumptions she makes about the assumptions she has in common with other agents.

An agent's cognitive environment, and thus her genre knowledge, grows out of her understanding—her assumptions—about the expectations of other persons around her for the appropriate or fitting response to a recurring, typified situation.

Conformity with genre conventions, including linguistic register, is a function of relevance, but the Writer's efforts may not be rewarded by success. The more accessible the agent's goals and the greater her commitment to them, the more cognitive effort she expends to develop assumptions about what formal aspects of an apparently genred text are essential to its genre status. This effort increases the likelihood that she will produce a text that conforms to her understanding of the genre. Of course, her text may not meet the expectations of a particular Reader for the genre. This is especially true where the Writer is an apprentice member of a disciplinary or professional community.

Most genres are not fixed things subject to categorical descriptions, just as the meaning of most words is not fixed or subject to categorical description.[3] Writers and Readers continually adjust their efforts to produce communicative performances that they think will optimize relevance. But there can be said to be consensus or near consensus on some formal characteristics of certain putative genres. For example, I speculate that an empirical study of practicing attorneys would permit the researcher to identify some central tendencies in their cognitive environments regarding the motion-to-dismiss memo; characterizing these central tendencies would be a way of identifying formal characteristics of the genre. In short, a genre can be seen as the central tendencies in the imputed cognitive environments of those who interact with the typified situation to which the genre is a response.

But I contend that the agent herself need never have experienced the typified situation to attempt a genre performance. This holds true of apprentice members of a discourse community as well. For example, if first-year law students come to a writing task with similar training about what formal characteristics are conventional in the motion-to-dismiss memorandum, and they are writing in response to the same (or at least very similar) social environment, they are writing in the same genre. Their

[3] Refer particularly to Section 2.4.1 at page 33 for the discussion of Sperber and Wilson's views regarding the process of interpretation by mutual adjustment.

work product may not exhibit the genre characteristics of motion-to-dismiss briefs written by experienced attorneys, but it is nevertheless the product of their shared genre knowledge—it is a genre.

Even experienced members of a professional community may have widely varying experience with a genre. I make the following observations based on anecdotal impressions from 15 years of law practice and some unsystematic review of memoranda filed in actual court cases. Imagine that two human agents—experienced attorneys—have identical "slices" of their cognitive environments that correspond to their understanding of what constitutes a memorandum in support of a motion to dismiss a case. The formal characteristics upon which these two agents agree can be seen as characteristics of the genre accessible to them; to the extent they are writing such a brief for a client, they also share the same goals.[4] The performances of another attorney attempting to enact this genre, however, may not conform to these formal characteristics in every respect. For example, the Writer may be an attorney of many years' practice who has never done litigation work, but is only now trying her hand at it. She may be a newly-minted lawyer who has some applicable genre knowledge but little practical experience. She may be responding to conventions arising in a different social environment; for example, perhaps she includes line-numbering in her brief, which is required by court rule in some jurisdictions but may not be required in the jurisdiction where she is filing the brief.

Privileging the genre performances of persons based on some arbitrary dividing line—whether they have graduated from law school, for example—reifies both the genres and the professional identities of the people who perform them. In reality, there is a wide variety or perhaps a continuum of assumptions held by persons engaged in professional activities that are not consistent with the central tendencies in those professions.

In the empirical study in this dissertation, the writers are students who have an unstable and rapidly evolving understanding of the conventions of legal writing in general and of the "genre" of the memorandum supporting or opposing a motion in particular. There is a high effort cost for them to keep accessible to themselves the assumptions they have about this genre. But as they are acutely aware of the stakes associated with

[4] We'll assume that they share an emotional state, too, though this question is taken up in greater detail in Section 7.4.

successful conformity, the accessibility and strength of their goals greatly increase the effect side of the relevance ratio.

6.6 CPR theory may help to explain findings in the present study

CPR theory may help to explain the findings of the empirical study in this dissertation: What they suggest is that gender differences, if they exist, are conventions that are vulnerable to being overwhelmed by other conventions (such as genre) when the Writer's cognitive environment makes the other conventions more relevant. Note the definitions offered earlier of genre and gender:

(13) Gender is a loosely and culturally defined set of social behaviors that are expected to make it possible to distinguish the two most common sexes from each other.

(14) A genre is a loosely and culturally defined set of communicative behaviors, usually formal conventions, a Speaker or Writer expects to have a particular effect or effects on a Hearer or Reader, based on assumptions about a typified situation in the Speaker's imputed cognitive environment.

At this point, it should be apparent that the definition of gender here is just a special case of the definition for genre. In each case, the accessible assumptions and goals of the agent should lead her to perform in a way that Hearers, Readers, or observers are expecting. In the case of gender, such performances are likely unconscious and habitual. They may be strategic in some cases (DeFrancisco et al., 2014), but are likely automatic in most. Thus in most cases, there would be no reason for an agent to vary from these conventional performances, unless there is an intervening goal that makes such a variation more relevant for the agent, and the agent has among her assumptions an alternative conventional path toward achieving that goal. The intervening goal in this empirical study was the high-stakes, graded assignment in the first-year legal writing course. The assumptions in the students' cognitive environment probably included representations about conventional language for legal writing. As a consequence, the genre of the legal brief may have overwhelmed any pattern of language differences associated with the students' genders (assuming, of course, they were there to begin with).

As a preliminary matter, this dissertation has taken it as a presumption that there generally *are* gender differences in language in informal contexts. Some of the studies described in Chapter 3 showed men and women conventionally engaged in differing communicative practices in contexts where they do not have a supervening goal to encourage variations from these conventions; these studies used large (or larger) data sets generally and generally showed there were differences in the writing of men and women (Argamon, Koppel, Fine, & Shimoni, 2003; Argamon, Koppel, Pennebaker, & Schler, 2007; Koppel et al., 2002; Rao et al., 2010; X. Yan & L. Yan, 2006). These studies were subject to methodological limitations—in particular a failure to make explicit their theories of gender and their methods for ascribing gender categories to participants and artifacts. Other studies reviewed in Chapter 3 showed that there were no gender differences when the participants were students training to write in professional or technical genres. Those studies had small sample sizes, so their ability to show the lack of differences is limited, and they suffered some of the same methodological problems as the studies that found gender differences. I have taken the position that tipping the scale here is folk belief: It is a commonly held belief that men and women are from different communicative cultures (DeFrancisco et al., 2014; Tannen, 2001). Thus, I have presumed that the students who came into this study wrote (and probably still do write) differently based on their genders in informal environments. This probably can and should be subjected to empirical scrutiny, but that task is beyond the scope of this dissertation, though I propose such an extension of this study in Section 7.4.

In general circumstances, CPR theory would predict men and women would exhibit those differences because of habit and because the differences are relevant to their interlocutors. These communicative habits arise over many years in individuals based on the conventional, typified situations in which they find themselves. Using habitual performances is low-cost for the agent. The agent will invest the cognitive effort into constructing a non-conforming performance only if she has a particular goal that is more accessible to her and to which she is more committed than signaling her gender. Agents in her audience will view a breach of a gender convention as less relevant—higher interpretation costs with no apparent payoff—unless the performance is situated with others that make the unconventional performance appear more relevant to the audience. Thus, in practice, unless an agent is challenging a gender convention or engaging in gender-play, she will prefer communicative performances that are consistent with gender expectations.

If the Speaker varies from habitual performances she imposes a search cost on herself to formulate an effective performance; she must consider her audience to produce an effective performance, one that imposes a processing cost on the Reader but justifies it by promising some additional effect. So a Speaker or Writer should be capable of suppressing gender differences (which is to say, she should be capable of departing from the genre in which gender consists). CPR theory predicts that a Writer will consciously vary from her habitual practices, which comes at a cost of effort, if the effect side of the relevance ratio makes it worth her while and if she has assumptions that will guide her.

The law school environment makes the conventions of the brief genre, including its linguistic register, relevant to the law students by making goals and assumptions related to it accessible. The first-year law school classroom provides an intensive initiation into a professional culture, and students are led to believe that their classroom performances, and the extent to which they satisfy their instructors' expectations, have high-stakes implications for their careers. Law students are pressured to conform to the linguistic registers of legal communication during their first year in law school. At matriculation, my intuition and anecdotal experiences suggest the linguistic registers and genres of legal writing are foreign to almost all the law students, regardless what their genders are. But they are presented with numerous examples of those registers in the court opinions that make up the majority of their reading during this time.

Assuming that students come to law school with habitually gendered communicative practices, they may nevertheless feel compelled to invest considerable cognitive effort in conforming to the law's registers and genres. In short, it is consistent with CPR theory that they would abandon (at least in their law-school writing) their previous communicative habits in order to achieve the expected benefits of doing so. As a result, we would expect that first-year law students' writing would not exhibit gender differences when they are attempting to write in a legal genre on an important graded assignment. We should observe gender-differences only where students fail to attend to genre and register conventions or where they are unable to discern them based on their training.

In short, CPR theory explains that apprentice writers in a new professional context will invest cognitive resources strategically to achieve their goals using assumptions they have developed about the professional conventions. The effort they will devote to such a task depends on the accessibility and strength of their goals. Their ability to conform depends on the assumptions available to them about the conventions. But note, as I have repeatedly done above, that this explanation of the findings in the empirical study in this dissertation is necessarily speculative, pending further qualitative research regarding the students' cognitive environments in courses such as these. I propose such an extension of this study in Section 7.4

6.7 Conclusion

This chapter has provided a reprise of CPR theory, recapitulating the contents of Chapter 2 in Section 6.2. In that section, I put CPR theory in the context of the law-school classrooms where the data for this study were collected. I also began in Section 6.3 a process of providing contextual information based on my own anecdotal experiences as a teacher in these classrooms and my intuitions arising from those experiences. I suggested that these assertions should be subjected to empirical inquiry; that exploration would likely take the form of qualitative interviews and observations.

In Section 6.4, I offered a CPR-theoretic account of gender. In brief, I conclude that gendered performances—including gender differences in informal language use—result from habitual practices engendered and reinforced by an agent's social environment. In any given case, whether it is makes sense to exhibit or suppress habitual gender performances will be a relevance-based assessment. In general, where there is no contextual reason for the agent to vary from habitual performances, we would expect habitual performances. For a Speaker to vary from such performances, she must invest cognitive effort, an investment she will make only if her accessible goals are strong enough to make the investment worthwhile.

I considered some aspects of a CPR-theoretic account of genre in Section 6.5. The Writers here were students who may have had an unstable and rapidly evolving understanding of the conventions of legal writing in general and of the "genre" of the memorandum supporting or opposing a motion in particular. If students were acutely aware, though, of the stakes associated with successful conformity, the accessibility and intensity of the students' goals would greatly have increased the effect side of the relevance ratio. This would have offset the high effort cost for them to keep accessible to themselves the assumptions they had about this genre and to consciously overcome any habitual linguistic performances—which are normally available at lower heuristic cost.

Thus, as I explained in Section 6.6, any abandonment by students in this study of gendered habitual linguistic performances (assuming they came to law school with them in the first place) in order to conform to the genre of the legal brief is consistent with CPR theory, because the students may have had accessible and strong goals that made it worthwhile for them to invest the cognitive effort in doing so.

Numerous questions remain: Though this study was intended to address methodological limitations of earlier studies, it is subject to limitations of its own. And though I contend that CPR theory provides a useful framework for explaining the results in this study, I have not yet made the case for its utility beyond that context. I will take up both these issues in Chapter 7.

Chapter 7

Conclusion

7.1 Introduction

CPR theory takes account of a Writer's goals and beliefs about the world and explains the efforts the Writer employs in finding, discovering, or inventing her communicative performances with the principle of relevance, which holds that she will expend effort in her writing choices that is commensurate with the accessibility and strength of the goals for which she writes and the effect she expects her writing to have on the cognition of the Readers(s). I have described this metaphorically as a ratio or fraction of cognitive effect to cognitive effort.

The study in this dissertation showed that apprentice writers may have abandoned presumed habitual stylistic differences that would normally vary with their genders in informal writing. I suggest they may have done so in response to a writing assignment in a genre, where they had been immersed for a few months in the genre's linguistic register, and where they likely perceived the stakes for conformity with genre and register conventions as being high. In short, I speculated that they may have invested significant cognitive effort to make assumptions about the genre and its register accessible to themselves and to conform their writing to it, because they expected a very significant effect: In that case, the task would have had high relevance in a CPR-theoretic sense.

This dissertation concludes with a summary of the limitations of the empirical study in it in Section 7.2; a discussion in Section 7.3 of the implications of this study and of potential implications and applications of CPR theory; and finally, in Section 7.4, suggestions of some questions for future study.

7.2 Limitations of this study

Before I claim in Section 7.3.1 that the empirical study in this dissertation is methodologically rigorous, I must acknowledge some limitations. Some of them are difficult to avoid in any study of this kind. Others result more from the fact that I did not know how I would interpret my findings until long after the data had been collected. As a consequence, these limitations will also prompt some of my suggestions for future research in Section 7.4. The limitations discussed here include the lack of longitudinal data or a pre-test of these participants, questions about participant self-selection and representativeness of these two law schools, concerns about statistical power and effect sizes in my findings, questions about the possible use of extensive quoted material in the students' samples, assertions that I have made in this dissertation that are grounded in my anecdotal knowledge rather than systematic observations, and questions about the process of feature reduction for machine-learning algorithms. I take up each of these concerns here.

Longitudinal data about participants

Ideally, this study would have included writing samples from study participants at the beginning of their first year in law school to see whether they really did come to law school with habitual stylistic differences that varied with their genders. As I have noted throughout this dissertation, previous studies have sometimes showed that authors' choices varied with their genders and sometimes showed that they did not. Rather than relying on a presumption grounded in the studies that were "positive" for gender differences and folk beliefs about gender differences, I would have preferred to measure this directly among the participants.

In fact, CPR makes some predictions that could only be tested with such a study. If students come to law school with gendered stylistic habits, we would expect those stylistic habits to show through in contexts where the relevance of the writing task was lesser. So, for example, if we asked law students to write a few paragraphs of "legal analysis" in response to an invitation to participate in a survey at the end of their first year, I anticipate that the students' goal of success at this task would be much less accessible and intense. In other words, performing a fragment of legal analysis in response to a researcher's offer of a gift card is not going to deliver enough effect, taking into account the students' goals, to make it worth the effort to suppress gendered stylistic habits. Contrast the high-stakes assignment studied in this project.

Participant self-selection and representativeness

The participants for this study were self-selected. They were not randomly selected from the population of first-year law students at Academy Law School and Lyceum College of Law. Approximately 36% (197 of 545) of the eligible students participated in the study. This is an encouraging response rate, but some other variable(s) common to the self-selected respondents could have affected the findings here. Given this type of research, which requires participants to provide gender self-identification, a different approach to the project (and the IRB) would likely be required to get a true random sample from the population of law students at these two schools.

What's more, these two schools, located in the Midwestern United States, might tend to have students who differ in some systematic way from law school students nationally. If students at these two schools do not matriculate with gendered stylistic habits, contrary to what some studies and folk beliefs suggest, then this study showed nothing. On the other hand, this would be a problem with any sample of students from law schools unless a pre-test were performed as suggested in the previous section.

Concerns about statistical power and effect sizes

This study consisted of a sample of papers from 193 gendered persons. (Though 197 writing samples were collected, four participants declined to identify their genders.) The test of statistical significance in this study was the Mann-Whitney U test (Utts & Heckard, 2006). Using this test, I assessed whether a measured difference between Gender F and Gender M authors in mean relative frequency of their use of some stylistic feature was likely the result of sampling. A difference in mean relative frequency measures is almost inevitable. A test of significance expresses the probability that the difference is due to chance. Thus, where the Mann-Whitney test provided a p-value less than 0.05, I could conclude that there was only a 1-in-20 probability that the difference was merely the result of sampling.

I did not, however, establish an a priori threshold for statistical *power* in my study. Post hoc power analyses are disfavored (Utts & Heckard, 2006, p. 474). Where an a priori power assessment is not performed, the *p*-value provides the best possible assessment of power. Power is a study's ability to claim that where no difference is visible in the sample, there is no underlying difference in the population that would have been visible if the sample size had been larger. Thus, at each place in this study where I say that there was "no difference" between texts written by Gender F and Gender M authors, the reader must consider whether a larger sample might have shown a difference. A sample size of 400 or 1,000 would serve to provide greater power. However, conducting a rigorously designed study with a much larger sample would have posed numerous practical problems.

A related question is statistical effect size. Assuming a difference in the relative frequency of use of a feature between Gender F and Gender M authors is statistically significant according to the Mann-Whitney test, one might reasonably ask whether the difference between the two samples reflects a *large* difference. Answering this question requires calculation of effect sizes. I have not calculated effect sizes for this study so far because of difficulties selecting the appropriate measure given that the samples here are in non-parametric distributions. Before seeking to publish the statistical findings from this study, I intend to undertake effect-size calculations and compare them to the findings in Argamon, Koppel, Fine, and Shimoni (2003), which did not provide effect sizes, but may have provided sufficient details to calculate them.

The possibility of extensive quoted material

I noted in Section 4.3.2 that I had excluded long "block quotes" from the writing samples before analyzing them but that I had left in places where students wove quoted material into their own writing. So, for example, the following sentence appears in paper 1102:

> The general rule under the Copyright Act is that a "work protected under this title vests initially in the author or authors of the work."

Such a sentence represents a hybrid of the student's language and the language of the quoted text because the student integrates her original composition with that of the quoted text. I did not attempt to identify the instances where students used such quotations frequently or where they constituted a large percentage of the student's paper. If lengthy and frequent quotations of this kind were very common in the data, they would serve to undermine the "single-author" nature of these samples and put my findings into question. Another pass through the samples to manually annotate quoted phrases would help to address that concern.

Data about the cognitive environment of participants

In several spots in this dissertation, particularly Sections 4.2 and 6.2, I made assertions about the material and cognitive environments of the participants in this study that were grounded not in my systematic observations, but in my anecdotal experience as a teacher for eight years in the legal writing classroom. I have attempted to note in each instance where I have done so. But as much of the explanation of the findings in this empirical study in Chapter 6 relies on these assertions, they should be subjected to empirical scrutiny.

Questions about MLA feature reduction

In Section 5.3.3, I discussed the fact that three different machine-learning algorithms (MLAs) worked best with three different reduced feature sets. Each of these MLAs—`SimpleLogistic`, `SMO`, and `NaiveBayes`—delivered classification results that were practically significant. One goal of the empirical study in this dissertation was to attempt to identify patterns of difference if the MLAs were successful in classifying these texts. In this case, however, the three MLAs used very different sets of features, subsets of the full 986-feature set that I created for this study, to achieve these results. I noted that the three algorithms made uses of different numbers of features, only 11 for `SimpleLogistic` and 43 for `NaiveBayes`; that the algorithms relied to different degrees on features that were statistically significant; and that they exhibited other variations.

My claim that those findings show no pattern of difference, though, is grounded in an understanding of the feature reduction process (described more fully in Section 4.4.2) that assumes the reduced feature sets to be a conclusive set of the features that allow a given MLA to perform its best. Without understanding the "mathematics" of the feature reduction algorithms, though, I am reluctant to rely to heavily on such an assumption. I would like to explore the feature reduction process further before attempting to publish these conclusions.

7.3 Implications, applications, and potential criticisms

The empirical study in this dissertation and the application of CPR theory to it have some fairly definite implications, which are taken up in Section 7.3.1. These include providing a demonstration of a methodologically rigorous gender-difference study; evidence for an "anti-essentialist" view of gender differences in communication; and insight into the process by which apprentice members of a profession adjust their communicative processes in response to their training. I hope and believe that CPR theory, once it is more fully developed, will lead to broader implications in rhetoric and in disciplinary, professional, and technical communication (DP&TC).[1] Examples discussed in Section 7.3.2 include nuancing rhetorical analysis and understanding the dynamics of rhetorical resistance. Finally, in Section 7.3.3, I consider some potential criticisms of CPR theory.

7.3.1 Implications of this study

The empirical study in this dissertation has several implications of value in writing studies and in DP&TC. First, this study has shown that it is possible to conduct methodologically rigorous gender-difference studies, using gender as a variable in an ethical fashion. Second, this rigorous study provides evidence for an "anti-essentialist" view of gender differences in communication. Third, this study may provide evidence for the process by which apprentice members of a profession adjust their communicative processes in response to their training.

[1] See page 2 for a definition and discussion of this term.

Methodologically rigorous gender-difference studies

This study has shown that methodologically rigorous gender-difference studies are possible in written communication. It provides guidance to future researchers on at least three fronts, which are largely related to the limitations of previous studies: gender construct, genre construct, and the single-author problem.

This study showed that gender can be used responsibly as a variable. I have argued that researchers using gender as a variable in studies of written communication owe to their participants and readers a thoughtful application of some gender construct. Readers are entitled to know what researchers intended to measure with a gender variable and how they measured it. I made three proposals originally in Section 3.4.2 that I have attempted to follow in this study: First, I proposed that researchers should not use gender as a variable in their studies unless a gender construct is necessary for answering their research questions. In this study, of course, gender was a central variable, so this forbearance was not possible. But I briefly described other survey research where gender could certainly have been collected as demographic variable (see, e.g., Larson et al. (In preparation), Pigozzi et al. (2014)), but where the researchers forbore from doing so on grounds that they had not developed a gender construct and gender was not essential for their study.

Second, I proposed that if a researcher uses gender as a variable, she has an ethical obligation to explain how she ascribed the variable. The least rigorous approach is for the researcher to acknowledge that she assigned the variable based on her personal impression after meeting or interacting with a participant. More rigorous approaches include using a gender role inventory instrument, like the BSRI, or obtaining participant self-identification for gender, as this study did. Finally, if the researcher collects participant self-identification, she should take care to offer options that represent a fuller range of gender identities. According to recent news reports, even Facebook has recognized the wisdom of moving away from the gender binary in user profiles (Whitney, 2015). In the present study, respect for the participants involved allowing participants to respond to a "gender" prompt with a free-form response.

This study also shows that communicative context can be taken into account when studying gender differences in language. CPR theory suggests that researchers looking for gender differences should also acknowledge the effect of participant history and habit interacting with communication task when assessing whether gender differences are present. Looking at writing produced by men and women who are writing informally with no common purpose (or genre) may result in finding gender differences; but those differences should not be taken as evidence of any essential difference arising from different communicative cultures. Rather, they likely represent habitual practices of the Writers, who for social reasons have previously written on topics and for purposes tied to their genders. This study has shown that there is no pattern of stylistic differences in the writing of men and women who have similar (albeit very brief) training and a similar goal for the writing task.

This study showed that it is possible to address the "single-author problem" in studies of authorial stylistics. It is difficult to say whether and to what extent previous studies were affected by researchers' lack of confirmation that texts they studied where written by single authors of the genders that they ascribed to the texts. Nevertheless, it is an essential problem when dealing with comparative stylistic studies of this kind. To say that "Authors of type A use stylistic feature Z more than authors of type B do" requires that the researcher make her best effort to ensure that she ascribes the A and B labels correctly.

Opposing gender-difference essentialism

I acknowledge the concerns of DeFrancisco et al. (2014) (and others) that studies of gender differences may essentialize differences between men and women. These authorities are likely concerned that gender-difference studies may reinforce folk beliefs that men and women are different in such ways that people will be prepared to accept gendered distribution of social roles on grounds of these essential differences. As I noted in Section 3.2, however, the most effective way to counter such essentialization is to conduct well–constructed gender-difference studies.

This study, for example, showed that men and women (or Gender F and Gender M authors) who were confronted with a similar writing goal after receiving similar training did not use stylistic features in ways that varied with their genders. In the context of professional writing at least, this study should lay to rest the idea that women and men come to the professional writing context with different communicative cultures that are difficult to change or might impede their professional acculturation.

Process of apprentice adaptation to professional discourse

In the same vein, this study may suggest ways in which men and women, whom we have presumed to come to the legal education environment with gendered communicative habits, acculturate to the professional context by abandoning gendered communicative practices. These observations, coupled with CPR theory, may tell us something about the ability of apprentice members of a professional discourse community to adapt to the new discourse conventions. Of course, we could learn more about that by looking at how the students' stylistic choices vary from those of practicing attorneys writing in the same genre, a matter that I take up in Section 7.4.

In addition to the implications of this particular empirical study and application to it of CPR theory, CPR theory has broader implications and potential applications.

7.3.2 Implications and applications of CPR theory

CPR theory has broad implications in understanding human communications in a variety of contexts. But I will identify here two implications and applications that seem obvious to me in the context of rhetoric and in disciplinary, professional, and technical communication (DP&TC): First, I believe CPR theory provides a basis to nuance rhetorical analysis to account for Speakers' and Hearers' cognitive environments as individuals. Without disregarding or disrespecting conceptions of social knowledge, CPR theory suggests that individual differences may play important roles in applying social knowledge. Second, CPR theory has implications for understanding rhetorical resistance and making it more effective. The Speaker who wishes to resist a social convention must be prepared to expend greater effort on invention in order to construct communicative performances that will appeal to the Hearers' expectations of relevance.

The broad implication of CPR theory is that understanding the production of utterances by Speakers and the interpretation of them by Hearers requires an appreciation of relevance. Speakers expecting to use unconventional communicative behavior or gender performances to achieve their goals should expect resistance form Hearers. Speakers can overcome resistance from Hearers by investing cognitive effort in understanding Hearers' goals and assumptions and using communicative performances to adjust them. These preparatory performances might be more-or-less conventional but can be seen as laying the groundwork for unconventional performances. Researchers in DP&TC need to account for these factors when studying workplace writing.

Consider an example: A manager in a company wishes to make small changes in the format of a report that circulates among other managers on a quarterly basis. The changes she wants to make affect the presentation of financial data in tables in the report but do not affect the executive summary that appears at the front of the report. She believes that the change she wishes to make will result in a clearer presentation of company fundamentals and that it will be more efficient (and therefore less expensive) for her staff to prepare. This Writer must first take account of the users of this report and their cognition regarding it. According to CPR theory, she can begin with the presumption that any change will be unwelcome. She should then account for the goals and assumptions of the managers who receive the report in determining how resistant they will be. If her assumption is that the other managers read the executive summary and refer to the tables only when they have questions, she may be less concerned about their resistance to the change. If on the other hand, her assumption is that one manager in particular would view the new tables as showing his department in a negative light, she will have inventive work to do, either to persuade him to accept the changes or to develop an alternative to her own original conception of the changes. If there are a great many managers who use the report, she might not know them all personally. Consequently, she may have to group them based on her understanding of their common characteristics, just as Aristotle did with a great many groups in the *Rhetoric* (Aristotle, 2007).

Of course, all this thinking about the cognitive environments of other managers comes with a cost for the Speaker. As she strategizes about changing the report, she must also consider whether the payoff to her, her department, or even the organization as a whole will be worth this effort. It is easy to see with such an example why organizational inertia often rules the day.

These principles and other concepts from CPR theory (which is explained in much more detail in Section 2.5) have other applications and implications, some of which I will summarize here.

Application to rhetorical analysis

CPR theory suggests ways to nuance rhetorical analysis. By considering the cognitive environment of each agent in a studied context, and by considering each agent's assumptions about other agents' cognitive environments (their imputed cognitive environments), the researcher may come to understand problems that go unstudied where the researcher is imposing a single system-model on the environment being studied. For example, models of rhetorical analysis grounded in activity systems theory (D. R. Russell, 1997), may tend to elide individual differences. I suggest that a researcher looking at an activity system with 15 subjects results in 16 mental representations of the activity system (one for each subject and one for the researcher), 16 (at least slightly) different conceptions of the activity system's objectives, maybe even 16 different views as to exactly who is a subject in the activity system. Any abstraction to observations about a social construct like an activity system, and indeed other social knowledge, is at most a center of gravity in a constellation of individual representations.

The CPR-theoretic approach would not revolutionize the writing classroom. Teachers and texts already urge student writers to consider the goals, needs, and values of audiences. But writing students may benefit greatly from understanding the relevance ratio: The Writer must be prepared to adjust her audience's goals and assumptions if she wants to depart from conventional communications in a technical or professional context. Otherwise, the Writer should be prepared to conform to conventions, even if she is hard-pressed to understand their purposes or is convinced that there is a better way.

Implications/applications in rhetorical resistance

CPR theory has implications for the concept of rhetorical resistance that can lead to applications. An agent who wishes to challenge conventional behavior or conventional framing of some social subject or object must consider the relevance ratio when predicting the reception her performance will generate. Consider two examples.

Challenging gender conventions comes with risks and cognitive costs. To position an action of resistance to be effective, an agent needs to think carefully about the risks of the resistance and how to situate it for the audience. Breaking gender conventions causes an immediate problem for the audience because the result is that the audience must engage in greater processing effort with no apparent offsetting positive effect. Thus the agent may wish to construct his/her resistance so as not to be threatening on other levels. An aggressive-sounding transvestite is less relevant to the audience than is an aggressive-sounding person or a transvestite, absent some reason for the audience to accept these performances. The agent can work to construct reasons, such as empathy etc., to win support for these performances. However, such strategic work for the agent comes at a heavy cognitive cost. We would expect the work of resistance to be hard, perhaps even exhausting, as one strategically assesses and adjusts the expectations of audiences. The agent who seeks to resist has to balance relevance for herself, to consider whether the effort required for effective resistance will be worth the distance the resistance carries her toward her goals.

Another example comes from the professional context: For 15 years or more, when giving presentations in business contexts, I have tried to use feminine pronouns as the default. When I talk about a generic lawyer or judge or real estate broker or consumer or teacher or doctor or nurse, I refer to "she," "her," etc. Initially, I adopted the practice in reaction to studies that showed that people used gendered pronouns to refer to professionals depending on the genders that predominated in those professions (masculine for doctors and judges and feminine for teachers and nurses, for example) and that the use of the pronouns resulted in Hearers visualizing persons of those genders in those roles. This is a goal that is readily accessible to me whenever I speak publicly because it has become my habit. I recognized (I think) or at least assumed those 15 years ago, that some folks would find the approach jarring or confusing. Nevertheless, I felt the cost to them would not be great enough for them to reject the content of what I was saying, and that my gender reframing might have positive (from my perspective) effects on their cognition. (Perhaps I am over-confident in my public-speaking skills.)

On one occasion, however, while giving a presentation on license law to real estate brokers in a rural Midwestern community, I received an overt challenge: An audience member, apparently elderly and male, exhibited body language during my presentation indicating that he was restless. I stopped to take questions before moving to another topic, and he aggressively asked, "Why do you keep saying 'she' instead of 'he'?" I believe what I did at that point was to assess the cognitive environments that I imputed to the audience. It was clear from the tone of his question that this man was bothered by my unconventional use of pronouns. But I also suspected that most of the rest of the audience did not really care much about it. However, given that I was speaking in a conservative community, I concluded that explaining the politics behind my stylistic choice would alienate the audience. At that point, my most accessible goal was to conclude the business presentation successfully; my political goal of reframing professional gender default was less accessible and strong to me at that time. I had choices about how to respond to his question, including making an overtly political statement about my strategy. Given my analysis of the situation, however, I responded simply by smiling, shrugging, and saying, "Why not?" I paused a moment to see if he could identify a reason he was willing to articulate, and then I moved on. I do not feel that I lost the audience at that point. Three audience members—apparently women—approached me after the presentation and thanked me particularly for that response.

Though I see CPR theory having implications for every aspect of human communication, these two examples, application of CPR theory to rhetorical analysis and implications of CPR theory to rhetorical resistance, suffice to show ways in which it may be valuable in the future. There are, however, several bases upon which CPR theory might be criticized.

7.3.3 Potential shortcomings of CPR theory

CPR theory might be criticized on several grounds. It might be attacked on grounds that it is reductionist, that it is scientistic or positivistic, or that is it a "theory of everything" and therefore a theory of nothing. I'll very briefly address these concerns here, but they will no doubt require a more full-throated rebuttal in the future.

CPR theory may be attacked as reductionist, in that it is looking at the "rhetorical situation" from the perspective of each agent, ignoring in some ways the social situation and notions of activity systems and the like. This is true, but then, theories grounded in social knowledge and activity systems are abstracting away from the cognition of individual agents—in short, they too are reductionist. All theories have to be reductionist in this sense. That is how they make knowledge.

CPR theory may be attacked as scientistic or positivistic. As I intend it as a quasi-foundational model for understanding human communication, it necessarily has roots in what I would call critical positivism or critical rationalism.[2] Though I value the ethnographic case study as a way of developing intuitions and theories about how things work, I value systematic investigation and quantitative methods as ways to confirm or enhance theories and to support cautious generalizations. Such knowledge can be made useful, and it has the benefit of being persuasive outside of our disciplinary boundaries.

CPR theory may be an ill-begotten "theory of everything" where communication theory is concerned. It may be too complicated, insufficiently grounded in the data upon which I hope it will rely, poorly characterized, just too boring to bother with, etc. If these things are true, then the efforts I make going forward to "package and sell it" to the disciplines of rhetoric and DP&TC will expose these flaws, and they may be fatal. Nevertheless, that will make knowledge, too, at least for me.

Because CPR theory is intended to explain all human communication from the individual agent's perspective, because it is meant to be grounded in an empirical research paradigm, and because it seeks to explain the cognition of those individual communicative agents, it may make several significant contributions to our knowledge about communication. These may include explaining why human agents select the communicative performances they do and how other communicative agents develop inferences about what they hear and read.

7.4 Questions for future research

Several questions for future research have already been suggested by the limitations of this study, discussed in Section 7.2, and among the implications of this study, discussed in Section 7.3. I will recap them here. I will follow them with other issues and questions that might productively be subjected to future research.

Issues arising from limitations and implications

As I've noted above, this study had several shortcomings, including a lack of longitudinal data about participants, potential issues with participant self-selection and representativeness, and questions about statistical power. These concerns can be addressed by gathering more samples from a broader group, which would help to address (but

[2] See Section 2.2 for a discussion of the metatheoretical underpinnings of CPR theory.

not eliminate) questions about self-selection and representativeness. In that process, it would make sense to perform pre- and post-tests, collecting writing samples from participants before their legal training begins and collecting writing samples from them at the end of their first years with lower stakes. For example, participants could be prompted to "write for 15 minutes about what you plan to do this summer." CPR theory would predict that such samples would show more evidence of gender differences than the formal memos written nearly contemporaneously with them.

At the same time, gathering richer, qualitative data from participants would offer many advantages. First, it would back up (or refute) claims that I made above about the students' cognitive environments, and especially about the accessibility and strength of their goals. Chapter 6 contains a great many hypotheses and speculations that should be subjected to empirical scrutiny. Analysis of such data could also extend the utility of CPR theory in the context of understanding apprentices' adaptation to professional discourse.

I anticipate that performing rhetorical analysis in other contexts using CPR theory might be difficult. For example, a CPR-theoretic analysis of Lincoln's second inaugural address would require the analyst to offer claims regarding Lincoln's cognitive environment and the cognitive environment(s) he imputed to his listeners, information that may be largely inaccessible to us now. The application of CPR theory would be much easier in studies of the kind discussed in the previous paragraph, where researchers could have ready access to the present impressions of the participants in the study, whether they are Writers or Readers. I believe that this application could potentially help those who attempt to engage in rhetorical resistance. I anticipate that folks who engage in such resistance regularly already instinctively understand the strategies they need to use. But I would like to see research into ways of educating people who are normally reluctant to abandon conformity to "try out" rhetorical resistance.

Other questions for future research

Several more questions for future research are top-of-mind for me, however, owing to questions I have received when discussing this project with others, to my interest in the process of typification, to my commitment that rhetoric and DP&TC must address emotions, and others.

Almost every time I have presented this research, at least one audience member has challenged me, asserting that the Gender F writers in my study have merely assimilated themselves to the masculine discourse of the law. The presumption is that the discourse of the law, because it was the exclusive domain of men for centuries and is still dominated by men at its upper echelons, must share characteristics with the habitual communicative practices of men. From that, it follows that law is a "native language" for men and a "foreign language" for women. At present, I have two responses to that question, each of which I would like to see tested in an empirical study.

First, I am not aware of any study that shows that legal writing is more like informal writing of men than informal writing of women. There are inferences that might be drawn, founded in folk beliefs about legal language and men's language and in earlier studies on gender difference. For example, the folk belief that men's writing seems more "informational" and less "involved," coupled with the intuition I have that legal writing is probably more informational than involved, might lead me to conclude that legal discourse is men's discourse. This is an empirical question, though, which no one appears to have answered with systematic study.

Second, my intuition from eight years of teaching first-year law students is that male students struggle to assimilate to these communicative conventions at least as much as females. If legal discursive conventions are masculine, I would expect male students, as native speakers, to have considerably less difficulty with them than female students. That has not been my experience. We could of course study this empirically, too. Both of these questions would make excellent studies. I would envision the former study as a quantitative analysis, and the latter using both qualitative and quantitative methods.

Another question important to me is how the typification of CPR theory takes place. I recounted in Section 3.5.4 the story told by Roger Schank in Schank and Abelson (1977) about his daughter making the purchase of a key chain part of the script of getting a new car. From the first experience that the toddler had of going with her dad to pick up a new car, she had latched onto one particular event (or subevent) that she made an important part of the story. In CPR-theoretic terms, there is a strong association between the car acquisition and the key chain acquisition in this example. Consequently, making the car purchase accessible would also make the key chain purchase accessible. In genre theory, it seems that we should similarly be able to account for how a first-time user of a text genre identifies those aspects of it that she believes (rightly or wrongly) to be essential components of it. In other words, how do users of text genres develop genre knowledge that allows them to make the correct strong associations between the context and text while avoiding extraneous associations? A study exploring this question would be important for teaching writing students how to acquire knowledge about new genres rather than just reproducing "toy" genres in the classroom.

A third significant question is of more theoretical interest. As I explained in Section 2.5, one component of every agent's cognitive environment is her emotional state. But I have left the question of emotional states largely out of the analysis and discussions in this dissertation. This disconnect is a function of two simple facts: First, my instinct is that emotions play a critical role in utterance production and interpretation. Second, in the time I have had to perform this analysis, I have not been able to puzzle out my own ideas about the role that emotions may be playing. In other words, approaching this question may not require a new empirical study; it may require a more nuanced analysis of the data in this study. Nevertheless, a qualitative study of Writers and Readers to assess their emotions when they are interacting with texts of this kind may prove useful.

A fourth question, and another of theoretical interest, relates to the epistemic possibilities of machine-learning algorithms in disciplinary, professional, and technical communication research. I've noted above that such tools enjoying use in practical applications in the field (Humpherys et al., 2011; McCart et al., 2013; Pakhomov, Hanson, et al., 2008; Pakhomov, Shah, et al., 2010). But Section 5.3 illustrates the difficulty of making sense or meaning from results of machine-learning trials. As these tools grow in their use for practical applications, it will become ever more important to theorize them and their operations.

A fifth opportunity for research, and the last I will discuss here, arises from the existence of the corpus created for this project. Perhaps these sample papers could be studied to learn whether other features in them vary significantly with author gender. For example, did the Gender F authors use more citations or headings than the Gender M authors? A related possibility is to compare these data with a corpus of texts written by experienced attorneys and actually filed in real cases. Understanding how the students' language use varies systematically from that of professional attorneys, if at all, could assist in focusing pedagogical efforts on teaching students to write more like the attorneys they want to become—or to rhetorically resist that kind of writing.

7.5 Conclusion

This dissertation began with a question:

> *How can cognitive pragmatic rhetorical (CPR) theory contribute to our understanding of rhetorical and disciplinary, professional, and technical communication theory and in particular to our theories of gender and genre performances?*

This dissertation has offered CPR theory as a theory of production and interpretation for written communication. It has shown how that theory may account for the lack of gendered stylistic choices in the writing of apprentice members of a professional community. And finally, it has suggested directions for future explorations, extensions, and applications of CPR theory.

The empirical study in this dissertation and the application of CPR theory to it have some fairly definite implications, including a demonstration of a methodologically rigorous gender-difference study; evidence for an “anti-essentialist” view of gender differences in communication; and some evidence for the process by which apprentice members of a profession adjust their communicative processes in response to their training. Once CPR theory is more fully developed, I want it to lead to broader implications in rhetoric and in DP&TC: It should help to explain the processes by which members of disciplinary or professional communities typify the communications in which they engage. And it should assist agents in understanding how most effectively to engage in rhetorical resistance.

References

Allen, J. (1994). Women and authority in business/technical communication scholarship: An analysis of writing... *Technical Communication Quarterly*, *3*(3), 271.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text*, *23*(3), 321–346.

Argamon, S., Koppel, M., Pennebaker, J. W., & Schler, J. (2007). Mining the blogosphere: Age, gender and the varieties of self-expression. *First Monday*, *12*(9). Retrieved September 27, 2010, from http://firstmonday.org/issues/issue12_9/argamon/index.html

Ariely, D. (2012, June). Essay mills – a coarse lesson on cheating. *Los Angeles Times*. Retrieved October 9, 2014, from http://articles.latimes.com/2012/jun/17/opinion/la-oe-ariely-cheating-20120617

Aristotle. (2007). *On rhetoric* (2nd) (G. A. Kennedy, Trans.). New York: Oxford University Press.

Ashley, K. & Bridewell, W. (2010). Emerging AI & law approaches to automating analysis and retrieval of electronically stored information in discovery proceedings. *Artificial Intelligence and Law*, *18*(4), 311–320.

Atwell, E. (n.d.). Automatic mapping among lexico-grammatical annotation models (AMALGAM): The University of Pennsylvania (Penn) treebank tag-set. University of Leeds, School of Computing. Retrieved February 1, 2014, from http://www.comp.leeds.ac.uk/amalgam/tagsets/upenn.html

Austin, J. L. (1975). *How to do things with words* (2nd ed.). Oxford; New York: Oxford University Press.

Bawarshi, A. S. & Reiff, M. J. (2010). *Genre: An introduction to history, theory, research, and pedagogy*. West Lafayette, IN: Parlor Press.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison: University of Wisconsin Press.

Bazerman, C. (2013). *A theory of literate action: Literate action Volume 2*. Fort Collins, CO: The WAC Clearinghouse.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*(2), 155–162.

Berkenkotter, C. & Huckin, T. N. (1994, November). *Genre knowledge in disciplinary communication: Cognition/culture/power*. Hillsdale, NJ: Lawrence Erlbaum.

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. Harlow, Essex, UK: Longman Group UK Ltd.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. New York: Cambridge University Press.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python* (1st ed.). Cambridge, MA: O'Reilly Media.

Bitzer, L. (1968). The rhetorical situation. *Philosophy & Rhetoric*, *1*(1), 1–14.

Blanchard-Fields, F., Suhrer-Roussel, L., & Hertzog, C. (1994). A confirmatory factor analysis of the Bem Sex Role Inventory: Old questions, new answers. *Sex Roles*, *30*(5-6), 423–457.

Board, I. R. (2004, October). Protecting human subjects guide. University of Minnesota. Retrieved from http://www.research.umn.edu/irb/download/Protecting%20Human%20Subjects%20Guide.pdf

Breuch, L.-A. K., Olson, A. M., & Frantz, A. (2002). Considering ethical issues in technical communication research. In L. J. Gurak & M. M. Lay (Eds.), *Research in technical communication* (pp. 1–22). Westport, CT: Praeger Publishers.

Brown, S. M. & Burnett, R. E. (2006, July). Women hardly talk. Really! Communication practices of women in undergraduate engineering classes. (T3F1–T3F9). San Juan, Puerto Rico: International Network for Engineering Education & Research. Retrieved October 11, 2013, from http://ineer.org/Events/ICEE2006/papers/3219.pdf

Burger, J., Henderson, J., Kim, G., & Zarrella, G. (2011, May). *Discriminating gender on Twitter*. MITRE Corporation. Bedford, MA. Retrieved from http://www.mitre.org/work/tech_papers/2011/11_0170/

Butler, J. (1997). *Excitable speech: A politics of the performative* (1st ed.). New York: Routledge.

Campbell, K. K. & Jamieson, K. (1978). Form and genre in rhetorical criticism: An introduction. In K. K. Campbell & K. H. Jamieson (Eds.), *Form and genre: Shaping rhetorical action* (pp. 9–32). Falls Church, VA: Speech Communication Association.

Carter, M. (2007, February). Ways of Knowing, Doing, and Writing in the Disciplines. *College Composition and Communication*, *58*(3), 385–418. doi:10.2307/20456952

Caulley, D. N. & Dowdy, I. (1986). Legal education as a model for the education of evaluators. *Educational Evaluation and Policy Analysis*, *8*(1), 63–75.

Cauthen, R. (2010). *Black letters: An ethnography of a beginning legal writing course.* Cresskill, NJ: Hampton Press.

Centers for Medicare and Medicaid Services. (2014, February). ICD-9-CM diagnosis and procedure codes: Abbreviated and full code titles. Downloadable versions of the diagnosis and procedure codes. Retrieved April 26, 2014, from http://www.cms.gov/Medicare/Coding/ICD9ProviderDiagnosticCodes/codes.html

Chapman, S. (2011). *Pragmatics*. New York: Palgrave Macmillan.

Churchland, P. (1989). *A neurocomputational perspective: The nature of mind and the structure of science*. Cambridge, MA: MIT Press.

Clark, A. (1997). *Being there: Putting brain, body, and world together again.* Cambridge, MA: MIT Press.

Clary, B. G. & Lysaght, P. (2006). *Successful legal analysis and writing: The fundamentals* (2nd ed.). St. Paul, MN: West Group.

Clary, B. G. & Lysaght, P. (2010). *Successful legal analysis and writing: The fundamentals* (3rd ed.). St. Paul, MN: Thomson Reuters.

Crystal, D. (1985). *A dictionary of linguistics and phonetics* (2nd ed.). London: Andre Deutsch.

Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, I., ... Peters, W. (2012, December). Developing language processing components with

GATE version 7 (a user guide). Retrieved January 1, 2013, from http://gate.ac.uk/sale/tao/split.html

Dascal, M. & Gross, A. G. (1999). The marriage of pragmatics and rhetoric. *Philosophy & Rhetoric*, *32*(2), 107–130.

DeFrancisco, V. P., Palczewski, C. H., & McGeough, D. D. (2014). *Gender in communication: A critical introduction* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Durack, K. (1997). Gender, Technology, and the History of Technical Communication. *Technical Communication Quarterly*, *6*(3), 249–60. Appears in Johnson-Eilola's Central Works.

Eaton, A., Brewer, P. E., Portewig, T. C., & Davidson, C. R. (2008). Examining editing in the workplace from the author's point of view. *Technical Communication*, *55*(2), 111–139.

Fausey, C. M. & Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, *17*(5), 644–650.

Figueroa, E. (1994). *Sociolinguistic metatheory*. Language & Communication Library. Tarrytown, NY: Elsevier Science.

Garrard, P., Maloney, L. M., Hodges, J. R., & Patterson, K. (2005). The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain*, *128*(2), 250–260. Retrieved from http://brain.oxfordjournals.org/content/128/2/250.abstract

Gibson, J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin Co.

Gigerenzer, G. & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–143.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Grush, R. (2001, February). The philosophy of cognitive science. Retrieved January 31, 2015, from http://mind.ucsd.edu/papers/phil_cogsci/phil_cogsci.html

Gundel, J. K. (2010, November). Reference and Accessibility from a Givenness Hierarchy Perspective. *International Review of Pragmatics*, *2*(2), 148–168. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=ufh&AN=55089150&site=ehost-live

Gutman, J. (1997). Means–end chains as goal hierarchies. *Psychology and Marketing*, *14*(6), 545–560.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009, July). The WEKA data mining software: An update. *SIGKDD Explorations*, *11*(1), 10–18. Retrieved February 9, 2014, from http://www.kdd.org/sites/default/files/issues/11-1-2009-07/p2V11n1.pdf

Hansen, S. (2004, August). Dear plagiarists: You get what you pay for. *The New York Times*. Retrieved October 9, 2014, from http://www.nytimes.com/2004/08/22/books/review/22HANSENL.html

Harris, R. A. (2007). Foreword to "Rudiments of cognitive rhetoric". *Rhetoric Society Quarterly*, *37*(4), 357–359.

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, *44*(1), 1–12.

Herring, S. C. & Paolillo, J. C. (2006). Gender and genre variation in weblogs. *Journal of Sociolinguistics*, *10*(4), 439–459.

Horn, L. R. (1992). Pragmatics, implicature, and presupposition. In W. Bright (Ed.), *International encyclopedia of linguistics* (Vol. 3, pp. 260–66). Oxford: Oxford University Press.

Hultgren, A. K. (2008). Reconstructing the sex dichotomy in language and gender research: Some advantages of using correlational sociolinguistics. In K. Harrington, L. Litosseliti, H. Sauntson, & J. Sunderland (Eds.), *Gender and language research methodologies* (pp. 29–42). New York: Palgrave Macmillan.

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *On Quantitative Methods for Detection of Financial Fraud*, *50*(3), 585–594.

Hutchison, K., Balota, D., Neely, J., Cortese, M., Cohen-Shikora, E., Tse, C.-S., ... Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods*, *45*(4), 1099–1114.

Jacobs, S. (2000). Rhetoric and dialectic from the standpoint of normative pragmatics. *Argumentation*, *14*, 261–286.

Jamieson, K. (1975). Antecedent genre as rhetorical constraint. *Quarterly Journal of Speech*, *61*(4), 406–415.

Janssen, A. & Murachver, T. (2004). The relationship between gender and topic in gender-preferential language use. *Written Communication*, *21*(4), 344–367.

Juarrero, A. (1999). *Dynamics in action: Intentional behavior as a complex system.* Cambridge, MA: MIT Press.

Jurafsky, D. & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Prentice Hall series in artificial intelligence. Upper Saddle River, NJ: Pearson Education.

Kahlenberg, R. D. (1999). *Broken contract: A memoir of Harvard Law School.* Amherst: University of Massachusetts Press.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review*, *93*(5), 1449–1475.

Kelly, A. R., Abbott, N. A., Harris, R. A., DiMarco, C., & Cheriton, D. R. (2010). Toward an ontology of rhetorical figures. In *SIGDOC '10 Proceedings of the 28th ACM International Conference on Design of Communication* (pp. 123–130). New York: ACM.

Kelly, A. R., McDougall, A., & Abbott, N. (2009). Rhetorical models for computational systems: An interdisciplinary approach to reusable, tailorable medical information. In *SIGDOC '09 Proceedings of the 27th ACM International Conference on Design of Communication* (pp. 155–162). New York: ACM.

Koppel, M., Argamon, S., & Shimoni, A. R. (2002). Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, *17*(4), 401–412.

Lakoff, G. & Johnson, M. (2003). *Metaphors we live by (with a new afterword).* Chicago: University of Chicago Press.

Lakoff, R. T. (2004). *Language and woman's place: Text and commentaries* (Revised and expanded ed.) (M. Bucholtz, Ed.). New York: Oxford University Press.

Larson, B. N., Pigozzi, L. M., & Lazaraton, A. (In preparation). *Instructional design for all: Universal design and students' linguistic and cultural backgrounds.*

Lay, M. M. (1989). Interpersonal conflict in collaborative writing: What we can learn from gender studies. *Journal of Business and Technical Communication*, *3*(2), 5–28.

Lazaraton, A. (2003). Evaluative criteria for qualitative research in applied linguistics: Whose criteria and whose research? *The Modern Language Journal*, *87*(1), 1–12.

Leech, G. (n.d.). A brief users' guide to the grammatical tagging of the British National Corpus. Retrieved April 19, 2014, from http://www.natcorp.ox.ac.uk/docs/gramtag.html

Liu, Y. & Zhu, C. (2011). Rhetoric as the antistrophos of pragmatics: Toward a "competition of cooperation" in the study of language use. *Journal of Pragmatics*, *43*(14), 3403–3415.

MacNealy, M. S. (1998). *Strategies for empirical research in writing*. Boston: Longman.

Maltz, D. N. & Borker, R. (1982). A cultural approach to male-female miscommunication. In J. J. Gumperz (Ed.), *Language and social identity* (pp. 196–216). Cambridge U.K.: Cambridge University Press.

Matthews, P. (2007). Register. Oxford: Oxford University Press. Retrieved from http://www.oxfordreference.com/10.1093/acref/9780199202720.001.0001/acref-9780199202720-e-2841

McCart, J. A., Berndt, D. J., Jarman, J., Finch, D. K., & Luther, S. L. (2013). Finding falls in ambulatory care clinical documents using statistical text mining. *Journal of the American Medical Informatics Association*, *20*(5), 906–914.

Mercier, H. & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*, 57–111.

Mertz, E. (2007). *The language of law school: Learning to "think like a lawyer"*. New York: Oxford University Press.

Miller, C. R. (1979, February). A humanistic rationale for technical writing. *College English*, *40*(6), 610–617. Appears in Johnson-Eilola's Central Works. ArticleType: primary_article / Full publication date: Feb., 1979 / Copyright © 1979 National Council of Teachers of English. Retrieved July 3, 2010, from http://www.jstor.org.floyd.lib.umn.edu/stable/375964

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech*, *70*(2), 151–167.

Miller, C. R. & Shepherd, D. (2004). Blogging as social action: A genre analysis of the weblog. In L. J. Gurak, S. Antonijevic, L. Johnson, C. Ratliff, & J. Reyman (Eds.), *Into the blogosphere: rhetoric, community, and culture of weblogs.* Retrieved January 24, 2011, from http://blog.lib.umn.edu/blogosphere/blogging_as_social_action_a_genre_analysis_of_the_weblog.html

Miller, C. R. & Shepherd, D. (2009). Questions for genre theory from the blogosphere. In J. Giltrow & D. Stein (Eds.), *Genres in the Internet: Issues in the theory of genre* (pp. 263–290). Philadelphia: John Benjamins.

Murphy, D. J. (2002). Surveys and questionnaires. In L. J. Gurak & M. M. Lay (Eds.), *Research in technical communication* (pp. 93–110). Westport, CT: Praeger Publishers.

Murray, M. & DeSanctis, C. H. (2009). *Legal writing and analysis.* New York: Thomson Reuters/Foundation Press.

Nerlich, B. & Clarke, D. D. (1996). *Language, action, and context: The early history of pragmatics in europe and america, 1780–1930.* Amsterdam Studies in the Theory and History of Linguistic Science. Amsterdam: John Benjamins.

Neumann, R. K. & Simon, S. (2011). *Legal writing* (2nd). New York: Aspen Publishers.

Newell, A. & Simon, H. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery*, *19*, 113–126.

Oates, L. C. & Enquist, A. (2006). *The legal writing handbook: analysis, research, and writing* (4th ed.). New York: Aspen Publishers.

La plume de ma tante (linguistics). (n.d.). Retrieved January 16, 2015, from http://en.wikipedia.org/wiki/La_plume_de_ma_tante_(linguistics)

Weka 3: Data Mining Software in Java. (n.d.). Retrieved January 1, 2013, from http://www.cs.waikato.ac.nz/ml/weka/

What is the BNC? (n.d.). Retrieved October 16, 2010, from http://www.natcorp.ox.ac.uk/corpus/index.xml

The Bluebook: A uniform system of citation (19th ed.). (2011). Cambridge, MA: The Harvard Law Review Association.

Osborn Jr., J. J. (2004, September). *The paper chase* (Special Anniversary Ed). Albany, N.Y: Whitston Publishing.

Oxford English Dictionary. (n.d.). "heuristic, n. and adj.". Retrieved from http://www.oed.com/view/Entry/86554?redirectedFrom=heuristic&

Pakhomov, S. V., Chacon, D., Wicklund, M., & Gundel, J. (2011). Computerized assessment of syntactic complexity in Alzheimer's disease: A case study of Iris Murdoch's writing. *Behavior Research Methods*, *43*(1), 136–144.

Pakhomov, S. V., Hanson, P. L., Bjornsen, S. S., & Smith, S. A. (2008). Automatic classification of foot examination findings using clinical notes and machine learning. *Journal of the American Medical Informatics Association*, *15*, 198–202.

Pakhomov, S. V., Shah, N. D., Hanson, P., Balasubramaniam, S. C., & Smith, S. A. (2010). Automated processing of electronic medical records is a reliable method of determining aspirin use in populations at risk for cardiovascular events. *Informatics in Primary Care*, *18*(2), 125–133.

Perelman, C. & Olbrechts-Tyteca, L. (1969, June). *The new rhetoric: A treatise on argumentation*. Notre Dame, IN: University of Notre Dame Press.

Peterman, M. L. (1997). The effects of concrete and abstract consumer goals on information processing. *Psychology & Marketing*, *14*(6), 561–583.

Pigozzi, L. M., Larson, B. N., & Lazaraton, A. (2014, August). *Writing for international students upper-division research progress report: Spring 2014 activities*. University of Minnesota-Twin Cities. Minneapolis.

Raign, K. R. & Sims, B. R. (1993). Gender, persuasion techniques, and collaboration. *Technical Communication Quarterly*, *2*(1), 89–104.

Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010, October). Classifying latent user attributes in Twitter. In *Proceedings of the 2nd International Workshop on Search and Mining User-generated Contents* (pp. 37–44). Toronto, ON, Canada: ACM.

Rehling, L. (1996). Writing together: Gender's effect on collaboration. *Journal of Technical Writing and Communication*, *26*(2), 163–176.

Russell, B. (1968). *The impact of science on society*. New York: AMS Press.

Russell, D. R. (1997). Rethinking genre in school and society: An activity theory analysis. *Written Communication*, *14*(4), 504–554.

Schank, R. C. & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum.

Schmedemann, D. A. & Kunz, C. L. (2007). *Synthesis: Legal reading, reasoning, and writing* (3rd ed.). New York: Aspen Publishers.

Schutz, A. (1964). *Sudies in social theory (Collected papers II)* (A. Brodersen, Ed.). Phaenomenologica. The Hague, Netherlands: Martinus Nijhoff.

Schutz, A. (1966). *Sudies in phenomenological philosophy (Collected papers III)* (I. Schutz, Ed.). Phaenomenologica. The Hague, Netherlands: Martinus Nijhoff.

Schutz, A. (1973). *The problem of social reality (Collected papers I)* (4th ed.) (M. Natanson, Ed.). Phaenomenologica. The Hague, Netherlands: Martinus Nijhoff.

Searle, J. R. (1970). *Speech acts: An essay in the philosophy of language*. London: Cambridge University Press.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge U.K.: Cambridge University Press.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*(1), 1–20.

Smeltzer, L. R. & Werbel, J. D. (1986). Gender differences in managerial communication: Fact or folk-linguistics? *Journal of Business Communication*, *23*(2), 41–50.

Smith, E. O. & Thompson, I. (2002). Feminist theory in technical communication making knowledge claims visible. *Journal of Business and Technical Communication*, *16*(4), 441–477.

Sotirin, P. (2000). 'All they do is bitch bitch bitch': Political and interactional features of women's officetalk. *Women & Language*, *23*(2), 19–25.

Sperber, D. & Cummins, S. (2007). Rudiments of cognitive rhetoric. *Rhetoric Society Quarterly*, *37*(4), 361–400.

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Wiley-Blackwell.

Spinuzzi, C. (1996). Pseudotransactionality, activity theory, and professional writing instruction. *Technical Communication Quarterly*, *5*(3), 295–308.

Sterkel, K. S. (1988). The relationship between gender and writing style in business communications. *Journal of Business Communication*, *25*(4), 17–38.

Straßheim, J. (2010). Relevance theories of communication: Alfred Schutz in dialogue with Sperber and Wilson. *Journal of Pragmatics*, *42*(5), 1412–1441.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge Applied Linguistics. Cambridge: Cambridge University Press.

Swales, J. M. (2004). *Research genres: Exploration and applications*. Cambridge Applied Linguistics. Cambridge: Cambridge University Press.

Tannen, D. (2001). *You just don't understand: Women and men in conversation* (1st ed.). William Morrow Paperbacks.

Tebeaux, E. (1990). Toward an understanding of gender differences in written business communications: A suggested perspective for future research. *Journal of Business and Technical Communication*, *4*(1), 25–43.

Thompson, I. (2004). Sex differences in technical communication: A perspective from social role theory. *Journal of Technical Writing and Communication*, *34*(3), 217–232.

Tong, A. & Klecun, E. (2004, June). Toward accommodating gender differences in multimedia communication. *IEEE Transactions on Professional Communication*, *47*(2), 118–129.

Turow, S. (2010). *One L*. New York: Farrar, Straus and Giroux.

Twenge, J. (1997). Changes in masculine and feminine traits over time: A meta-analysis. *Sex Roles*, *36*(5-6), 305–325.

Utts, J. M. & Heckard, R. F. (2006). *Statistical ideas and methods*. Belmont, CA: Thomson Brooks/Cole.

van Eemeren, F. H. & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge U.K.: Cambridge University Press.

Whitney, L. (2015, February). Facebook now lets you describe your gender. *C—NET*. Retrieved March 20, 2015, from http://www.cnet.com/news/facebook-now-lets-you-describe-your-gender-in-your-profile/

Wilson, D. & Sperber, D. (2006). Relevance theory. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 607–632). Hoboken, NJ: Wiley-Blackwell.

Wilson, D. & Sperber, D. (Eds.). (2012). *Meaning and relevance*. Cambridge U.K.: Cambridge University Press.

Witten, I. H., Eibe, F., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Margan Kaufmann series in data management systems. Burlington, MA: Elsevier/Morgan Kaufmann.

Wolfe, J. & Alexander, K. P. (2005). The computer expert in mixed-gendered collaborative writing groups. *Journal of Business and Technical Communication*, *19*(2), 135–170.

Wolfe, J. & Powell, E. (2006). Gender and expressions of dissatisfaction: A study of complaining in mixed-gendered student work groups. *Women & Language*, *29*(2), 13–20.

Wolfe, J. & Powell, E. (2009). Biases in interpersonal communication: How engineering students perceive gender typical speech acts in teamwork. *Journal of Engineering Education*, *98*(1), 5–16.

Wu, H. & Shapiro, J. L. (2006). Does overfitting affect performance in estimation of distribution algorithms. In *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation* (pp. 433–434). Seattle, Washington, USA: ACM.

Yan, X. & Yan, L. (2006). Gender classification of weblog authors. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs* (pp. 228–230).

Yus, F. (2015, January). Relevance theory online bibliographic service. Retrieved January 31, 2015, from http://personal.ua.es/francisco.yus/rt.html

ZachWeinberg. (n.d.). SVM separating hyperplanes (SVG). licensed under Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons. Retrieved from http://commons.wikimedia.org/wiki/File:Svm_separating_hyperplanes_(SVG).svg#mediaviewer/File:Svm_separating_hyperplanes_(SVG).svg

Zheng, J. & Tang, Y. (2005, June). One generalization of the Naive Bayes to fuzzy sets and the design of fuzzy Naive Bayes classifer. In J. Mira & J. R. Álvarez (Eds.), *Artificial intelligence and knowledge engineering applications: A bioinspired approach* (pp. 281–90). Springer.

Appendix A

Project materials deposited in the University of Minnesota Digital Conservancy

A.1 Introduction

This appendix includes a discussion of the materials relating to this dissertation project that are available in the University of Minnesota Digital Conservancy. My purpose in making these materials available here is to expose my research materials for use by other researchers, whether they seek to confirm or disconfirm my findings or to extend the research I have performed here.

Beginning in May 2017, it should be possible to locate this dissertation and the materials described in this appendix by using the following search phrase in your favorite search engine: `University of Minnesota Brian Larson Gender Genre dissertation`. Once you receive search results, select the link where the URL begins `conservancy.umn.edu`. From the resulting abstract page, you should be able to download a single .ZIP file that contains the PDF of this dissertation and a folder titled "Appendix A materials." This appendix describes the Appendix A materials.

In the event that you have questions regarding these materials, you may reach the author at his research email address: Larson@Rhetoricked.com.

A.2 Contents overview

The contents of the Appendix A materials folder consist of four subfolders:

1. Raw papers de-identified: This folder contains 197 student writing samples, with all personally identifying information removed. The files are in Microsoft Word format, with the name of the file corresponding to each papers unique ID. Section 4.3.1 describes how the unique IDs are assigned. References in this dissertation to paper numbers are to these IDs.

2. Survey responses: This folder contains a single Microsoft Excel worksheet. Its contents are described further below.

3. Papers in XML: This folder contains an XML file for each student's paper. Their structure is described further below.

4. Feature tables: This folder contains a single table in WEKA's .ARFF format with the values for all the features for each of the student papers, and a single Microsoft Excel worksheet with the same values as well as additional values discussed in this dissertation. They are described further below.

A.3 Survey responses

The Microsoft Excel worksheet with the survey responses has the name "Master survey response and notes.xlsx" and has the following structure. First, each row represents one sample paper, except that the first row has column headings. The first column, titled "UniqueID," provides a unique ID for each paper in the study. Section 4.3.1 describes how the unique IDs are assigned. References in this dissertation to paper numbers are to these IDs. The remainder of the columns correspond to the survey questions described in Appendix E in a fairly obvious way. Note that the "SI_Gender" column provides the student's free-form response to the gender self-identification question. See Section 4.2.3 for a discussion of how the "Gender F" and "Gender M" labels were applied to these papers based on the SI_Gender response.

A.4 Papers in XML

There is one XML file for each sample paper. The XML files were originally created by means of an export from GATE. I then used Python to layer in additional information and to create excerpts of the papers for analysis within the XML file for each paper. (See Chapter 4 and Appendix G for more details on these processes.)

The significant sections of the XML document in each case are these:

- The `TextWithNodes` section contains the entire text of the student's paper with nodes identified that are used in the XML offset below. Each node represents the starting point or ending point of an annotation.

- An `AnnotationSet` section represents my annotations or those of a research assistant on the document in GATE. Each annotation has a `StartNode` and `EndNode` that correspond to the nodes in the `TextWithNodes` section of the XML document. Annotations correspond to the coding guide described in Appendix G. There is more than one annotation set on some papers where both the research assistant and I coded the paper.

- The GG section represents information from outside the paper added to the XML file. That includes the `PaperNum` value, which is equal to the paper's unique ID number. It also includes information from the questionnaire to which students responded, as well as some additional data I coded based on a qualitative review of the papers.

- The `CleanText` section has three subsections:

  - The `CleanFull` section is the full text of each memo as analyzed for this dissertation, but with the citations, headings, and block quotes stripped out. (See Appendix G for details.)
  - The `CleanNonFact` section is the full text of each memo, but with the fact section of the memorandum excluded and with the citations, headings, and block quotes stripped out.
  - The `CleanFact` section is just the fact section of each memo with the citations, headings, and block quotes stripped out.

A.5 Feature tables

There are two files in the "Feature tables" folder. The first is a Microsoft Excel file titled "All features.xlsx." This file contains all the feature values calculated for the analyses in this dissertation. The first row contains column headings, and each subsequent row provides data relating to one of the student papers.

Here is a summary of the column contents:

- Columns with names beginning with "A_" provide summary values for the papers.
  - "A_papernum" provides the unique ID of each paper.
  - "A_gender" provides the gender, "1" for Gender F and "0" for Gender M.
  - "A_tokens" provides a count of the total number of tokens in the paper.
  - "A_sents" provides a count of the total number of sentences in the paper.
  - "A_sentLen" is a calculated value approximating mean sentence length and equaling A_tokens/A_sents.
- Columns with names beginning "AK_" provide calculated values used in the comparison of this study to the Argamon/Koppel 02/03 study.
- Columns with names beginning "Bi_" provide relative frequencies for the bigrams identified.
- Columns with names beginning "F_" provide relative frequencies for function words.
- Columns with names beginning "POS_" provide relative frequencies for parts of speech.
- Columns with names beginning "Tri_" provide relative frequencies for the trigrams identified.

The second file titled "Completetexts.arff" contains the same information as the Excel file but is formatted for use with the WEKA machine-learning platform.

Appendix B

Part-of-speech tags

This appendix identifies the part-of-speech tags used throughout this study, based on the Penn Treebank tag-set. The process for part-of-speech tagging is is described briefly in Chapter 4 and more fully in Appendix G. (Note particularly the discussion in Appendix G.2.) This table consists of a single table which copies Atwell (n.d.) nearly verbatim. I have abbreviated some of the examples. I have also changed some of the tag names to make it easier to read other tables in this dissertation. So, for example, instead of [,] as the tag for a comma, I have used [Comma]. I've indicated this in the table below by placing the tag name I used in parentheses next to the Penn Treebank tag. There were also several tags not used in this project to avoid confounding NLP tools in the context of legal text. These tags are marked with "not used" in parentheses after the name of the tag.

Table B.1: **Part-of-speech tags used in this study (Penn Treebank)**

Tag	Description	Examples
, (Comma)	comma	,
: (Colon)	colon or ellipsis	: ; ...
. (Period)	sentence terminator	. ! ?
$	dollar	$ -$ –$ A$ C$ HK$ M$ NZ$ S$ U.S.$ US$
" (OQuote)	opening quotation mark	' "
" (CQuote)	closing quotation mark	' "
((not used)	opening parenthesis	([
) (not used)	closing parenthesis	)]
– (not used)	dash	–

Table B.1: Part-of-speech tags... (continued)

Tag	Description	Examples
CC	conjunction, coordinating	& 'n and both but either et for less minus neither nor or plus so therefore times v. versus vs. whether yet
CD	numeral, cardinal	mid-1890 nine-thirty forty-two one-tenth ten million 0.5 one forty-seven 1987 twenty '79 zero two 78-degrees eighty-four IX '60s .025 fifteen 271,124 dozen quintillion DM2,000 ...
DT	determiner	all an another any both del each either every half la many much nary neither no some such that the them these this those
EX	existential there	there
FW	foreign word	gemeinschaft hund ich jeux habeas Haementeria Herr K'ang-si vous lutihaw alai je jour objets salutaris fille quibusdam pas trop Monte terram fiche oui corporis ...
IN	preposition or conjunction, subordinating	astride among upon whether out inside pro despite on by throughout below within for towards near behind atop around if like until below next into if beside ...
JJ	adjective or numeral, ordinal	third ill-mannered pre-war regrettable oiled calamitous first separable ectoplasmic battery-powered participatory fourth still-to-be-named multilingual multi-disciplinary ...
JJR	adjective, comparative	bleaker braver breezier briefer brighter brisker broader bumper busier calmer cheaper choosier cleaner clearer closer colder commoner costlier cozier creamier crunchier cuter ...
JJS	adjective, superlative	calmest cheapest choicest classiest cleanest clearest closest commonest corniest costliest crassest creepiest crudest cutest darkest deadliest dearest deepest densest dinkiest ...
LS	list item marker	A A. B B. C C. D E F First G H I J K One SP-44001 SP-44002 SP-44005 SP-44007 Second Third Three Two * a b c d first five four one six three two
MD	modal auxiliary	can cannot could couldn't dare may might must need ought shall should shouldn't will would

Table B.1: Part-of-speech tags. . . (continued)

Tag	Description	Examples
NN	noun, common, singular or mass	common-carrier cabbage knuckle-duster Casino afghan shed thermostat investment slide humour falloff slick wind hyena override subhumanity machinist ...
NNP	noun, proper, singular	Motown Venneboerger Czestochwa Ranzer Conchita Trumplane Christos Oceanside Escobar Kreisler Sawyer Cougar Yvette Ervin ODI Darryl CTCA Shannon A.K.C. Meltex Liverpool ...
NNPS	noun, proper, plural	Americans Americas Amharas Amityvilles Amusements Anarcho-Syndicalists Andalusians Andes Andruses Angels Animals Anthony Antilles Antiques Apache Apaches Apocrypha ...
NNS	noun, common, plural	undergraduates scotches bric-a-brac products bodyguards facets coasts divestitures storehouses designs clubs fragrances averages subjectivists apprehensions muses factory-jobs ...
PDT	pre-determiner	all both half many quite such sure this
POS	genitive marker	' 's
PRP	pronoun, personal	hers herself him himself hisself it itself me myself one oneself ours ourselves ownself self she thee theirs them themselves they thou thy us
PRP$	pronoun, possessive	her his mine my our ours their thy your
RB	adverb	occasionally unabatingly maddeningly adventurously professedly stirringly prominently technologically magisterially predominately swiftly fiscally pitilessly ...
RBR	adverb, comparative	further gloomier grander graver greater grimmer harder harsher healthier heavier higher however larger later leaner lengthier less-perfectly lesser lonelier longer louder lower more ...
RBS	adverb, superlative	best biggest bluntest earliest farthest first furthest hardest heartiest highest largest least less most nearest second tightest worst

Table B.1: Part-of-speech tags... (continued)

Tag	Description	Examples
RP	particle	aboard about across along apart around aside at away back before behind by crop down ever fast for forth from go high i.e. in into just later low more off on open out over per pie raising start teeth that through under unto up up-pp upon whole with you
SYM (not used)	symbol	% & ' " ".)). * + ,. ¡ = ¿ @ A[fj] U.S U.S.S.R
TO	"to" as preposition or infinitive marker	to
UH	interjection	Goodbye Goody Gosh Wow Jeepers Jee-sus Hubba Hey Kee-reist Oops amen huh howdy uh dammit whammo shucks heck anyways whodunnit honey golly man baby diddle hush sonuvabitch ...
VB	verb, base form	ask assemble assess assign assume atone attention avoid bake balkanize bank begin behold believe bend benefit bevel beware bless boil bomb boost brace break bring broil brush build ...
VBD	verb, past tense	dipped pleaded swiped regummed soaked tidied convened halted registered cushioned exacted snubbed strode aimed adopted belied figgered speculated wore appreciated contemplated ...
VBG	verb, present participle or gerund	telegraphing stirring focusing angering judging stalling lactating hankerin' alleging veering capping approaching traveling besieging encrypting interrupting erasing wincing ...
VBN	verb, past participle	multihulled dilapidated aerosolized chaired languished panelized used experimented flourished imitated reunifed factored condensed sheared unsettled primed dubbed desired ...
VBP	verb, present tense, not 3rd person singular	predominate wrap resort sue twist spill cure lengthen brush terminate appear tend stray glisten obtain comprise detest tease attract emphasize mold postpone sever return wag ...

Table B.1: Part-of-speech tags... (continued)

Tag	Description	Examples
VBZ	verb, present tense, 3rd person singular	bases reconstructs marks mixes displeases seals carps weaves snatches slumps stretches authorizes smolders pictures emerges stockpiles seduces fizzes uses bolsters slaps speaks pleads ...
WDT	WH-determiner	that what whatever which whichever
WP	WH-pronoun	that what whatever whatsoever which who whom whosoever
WP$	WH-pronoun, possessive	whose
WRB	Wh-adverb	how however whence whenever where whereby whereever wherein whereof why

Appendix C

Function words used in the Argamon/Koppel 02/03 study and in the present study

This appendix consists of the list of function words used in this study. The process is described briefly in Chapter 4. The list was provided by Dr. Moshe Koppel in response to my email inquiry. The list(s) originally associated with Argamon, Koppel, Fine, and Shimoni (2003) and Koppel et al. (2002) were at a URL that was no longer functioning as of late 2013. Dr. Koppel provided the list below, expressing a hope that it was the list used in the earlier studies.

Note that function words marked with a dagger (†) did not appear in the corpus in this study, and they are thus not listed in the frequency table for function words in Appendix I (Table I.2). Many of the function words not appearing in the present study were archaic (e.g., *dost, doth, oft, thee, thy, tis*). Many others were numbers that did not correspond to the subject matter in these briefs (e.g., *billion, billionth, hundredth, ninetieth, thirteenth*).

'd (as in *he'd*)	's (as in *she's*)	above	actually
'll (as in *he'll*)	've (as in *they've*)	according	after
'm (as in *I'm*)	a	accordingly	afterward
're (as in *we're*)	about	actual	afterwards

again
against
ago
ah †
ai †
all
almost
along
already
also
although
always
am
among
an
and
another
any
anybody
anyone
anything
anywhere
are
around
art
as
aside
at
away
ay †
back

be
bear
because
been
before
being
below
beneath †
beside †
besides
better
between
beyond
bid †
billion †
billionth †
both
bring
but
by
ca[1]
came
can
cannot †
canst †
certain
certainly
come
comes
consequently
could

couldst †
dear †
definite
definitely
despite
did
do
does
doing
done
dost †
doth †
doubtful
doubtfully †
down
due
during
e.g. †
each
earlier
early
eight
eighteen
eighteenth
eighth
eighthly †
eightieth †
eighty
either
eleven
eleventh

else
enough
enter
ere †
erst †
even
eventually
ever
every
everybody †
everyone
everything
everywhere †
example
except
exeunt †
exit
fact
fair
far
farewell †
few
fewer
fifteen
fifteenth
fifth
fifthly †
fiftieth †
fifty
finally
first

[1] Contraction of ‘can’ as in *can’t*.

firstly
five
for
forever
forgo
forth
fortieth †
forty
four
fourteen
fourteenth
fourth
fourthly †
from
furthermore
generally
get
gets †
getting
give
go
good
got
had
has
hast †
hath †
have
having
he
hence
her
here
hers
herself
him
himself
his
hither †
ho †
how
however
hundred
hundredth †
I
if
in
indeed
instance
instead
into
is
it
its
itself
last
lastly
later
less
let
like
likely
many
matter
may
maybe
me
might
million
millionth †
mine
more
moreover
most
much
must
my
myself
n't[2]
nay
near
nearby
nearly
neither
never
nevertheless
next
nine
nineteen
nineteenth †
ninetieth †
ninety †
ninth
ninthly †
no
nobody
none
noone †
nor
not
nothing
now
nowhere
o
occasionally
of
off
oft †
often
oh †
on
once
one
only
or
order
other
others
ought
our
ours
ourselves
out
over

[2] Contraction of 'not' as in *can't*.

perhaps
possible
possibly
presumable †
presumably
previous
previously
prior
probably
quite
rare
rarely
rather
result
resulting
round
said
same
say
second
secondly
seldom
seven
seventeen
seventeenth †
seventh
seventhly †
seventieth †
seventy
sha[3] †
shall
shalt †
she
should
shouldst †
similarly
since
six
sixteen
sixteenth †
sixth †
sixthly †
sixtieth †
sixty
so
soever †
some
somebody
someone
something
sometimes
somewhere
soon
still
subsequently
such
sure
tell
ten
tenth †
tenthly †
than
that
the
thee †
their
theirs
them
themselves
then
thence †
there
therefore
these
they
thine †
third
thirdly
thirteen
thirteenth †
thirtieth †
thirty
this
thither †
those
thou †
though
thousand
thousandth †
three
thrice †
through
thus
thy †
till †
tis †
to
today
tomorrow
too
towards
twas †
twelfth †
twelve
twentieth
twenty
twice
twill †
two
under
undergo
underneath †
undoubtedly
unless
unlikely
until
unto
unusual
unusually
up
upon
us
very
was

[3] Beginning of *shan't*.

wast †
way
we
welcome †
well
were
what
whatever
when
whence †
where
whereas
wherefore
whether
which
while
whiles †
whither †
who
whoever
whom
whose
why
wil †
will
wilst †
wilt †
with
within
without
would
wouldst †
ye †
yes
yesterday
yet
you
your
yours †
yourself †
yourselves †

Appendix D

Research information form for study participants

The *Information Sheet for Research* distributed to students with the invitation to participate in this study appears on the following two pages.

INFORMATION SHEET FOR RESEARCH
Analysis of Law Student Writing Assignments

You are invited to be in a research study of law student writing assignments. You were selected as a possible participant because you are a first-year law student enrolled in a legal writing course. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by: Brian N. Larson, J.D., Writing Studies Department, University of Minnesota ([email]).

Procedures:

If you agree to be in this study, we would ask you to do the following things:
* Indicate your consent below to proceed to a brief online survey.
* When the survey begins, you will be asked to upload a copy of the final version of your major spring writing assignment in law school.
* The survey itself will ask you some demographic questions and will take no longer than 12 minutes to complete.
* When you have completed the survey, you will be able to provide your email address, which is where the researcher will send your code for a $15 Amazon gift certificate, provided in gratitude for your willingness to participate in the study.
* After any identifying marks (your name, address, phone number, email address, etc.) are removed from your writing sample, it may be published as part of a database of student papers that other researchers may use for other projects. By consenting below, you are consenting to the ongoing use of your writing sample by other researchers.

Confidentiality:

The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records. After any identifying marks (your name, address, phone number, email address, etc.) are removed from your writing sample, it may be published as part of a database of student papers that other researchers may use for other projects. By consenting below, you are consenting to the ongoing use of your writing sample by other researchers.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota or [Name of law school]. If you decide to participate, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions:

The researcher(s) conducting this study is (are): Brian N. Larson and his supervisor, Mary Lay Schuster. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact them at Department of Writing Studies, University of Minnesota, [phone], [email]. Larson's advisor, Mary Lay Schuster, is available at the Department of Writing Studies, University of Minnesota, [email and phone].

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Appendix E

Student survey instrument

The survey questionnaire instrument to which participants in the empirical study responded appears here. For an description of its design and administration, see Sections 4.2.3 and 4.3.

Student survey

Please upload your year-end legal writing assignment by clicking on the link at the [right/left/above]. Please answer the following questions. You are not required to answer any of these questions, but your answers may assist in making the research results more useful.

Age: [multiple choice consisting of following options]

- Under 18
- 18-24
- 25-33
- 34-45
- 46+

Gender: [blank box permitting a response not exceeding 10 characters]

Highest level of education you have completed: [multiple choice consisting of following options]

- Bachelors degree (U.S. institution)
- Bachelors degree or equivalent (Institution outside U.S.)
- Law degree (Institution outside U.S.)
- Master's Degree, post-baccalaureate professional degree, or equivalent
- PhD or equivalent
- Other [blank box permitting response]

Before your current legal writing/research course, when is the last time you took a course that you would describe as a "writing course": [multiple choice consisting of following options]

- I have never taken any other writing course.
- I took a writing course in a post-baccalaureate degree-granting program.
- I took a writing course as an upper-level undergraduate in the U.S.
- I took a writing course as a lower-level or freshman undergraduate in the U.S.
- I took a writing course as a student at a university outside the U.S.
- I took a writing course in secondary school (high school, for U.S. students)
- Other [blank box permitting response]

Describe how you learned English:

- I learned English in the U.S. as my first language
- I learned English in the U.S. as my second (or subsequent) language
- I learned English outside the U.S. as my first language
- I learned English outside the U.S. as my second (or subsequent language)

What is the section number of your legal writing class in law school? [blank space permitting response]

What is the last name of your legal writing professor or instructor? [blank space permitting response]

Email address: (You must answer this question in order to receive your $15 Amazon gift card) [blank box permitting student to enter email address]

Appendix F

Demographics of student participants

The following pages provide a description of demographics of the students participating in the empirical study (n=197; note that four respondents did not provide an answer to the gender self-identification question, and their papers were therefore not included in the analyses in this dissertation). See Appendix E for the complete survey instrument. The following demographic categories are provided here:

- Gender
- Age
- Education
- Last writing course
- Where participant learned English

The responses of participants to these questions are stored with individual sample papers in the XML data files described in Appendix A.

F.1 Gender self-identification

See Table 4.1 at page 112 for details of participants' responses to the gender self-identification question.

F.2 Other demographics

Table F.1: **Demographics: Participant age**

Category	Number of responses
18-24	109
25-33	78
34-45	7
46+	3
Total	197

Table F.2: **Demographics: Participant education**

Category	Number of responses
Bachelors degree (U.S. institution)	165
Bachelors degree or equivalent (Institution outside U.S.)	14
Master's Degree; post-baccalaureate prof. degree; equivalent	17
Law degress (Institution outside U.S.)	0
Ph.D. or equivalent	0
Not answered/decline to answer	1
Total	197

Table F.3: **Demographics: Participant's last writing course**

Category	Number of responses
I took a writing course. . .	
. . . as a lower-level or freshman undergraduate in the U.S.	57
. . . as a student at a university outside the U.S.	1
. . . as an upper-level undergraduate in the U.S.	93
. . . in a post-baccalaureate degree-granting program.	17
. . . in secondary school (high school, for U.S. students)	10
I have never taken any other writing course.	13
"Other" free-form responses	
'Creative writing course for law school applications'	1
'English major'	1
'I have taken writing intensive courses, but not courses specifically focused on teaching writing skills'	1
'I took a independent research class'	1
'Law school 1L'	1
Not answered/decline to answer	1
Total	197

Table F.4: **Demographics: Where participant learned English**

Category	Number of responses
I learned English. . .	
. . . in the U.S. as my first language	177
. . . in the U.S. as my second (or subsequent) language	9
. . . outside the U.S. as my first language	3
. . . outside the U.S. as my second (or subsequent language)	5
Not answered/decline to answer	3
Total	197

Appendix G

Data preparation

This appendix explains in detail the following processes applied to the corpus of data collected for the empirical study described in Chapter 4. From this corpus, I then needed to create abstractions, for each paper an *instance* consisting of *attributes* that the machine learning algorithms (MLAs) could process.

This process involved the following steps:

1. Manually annotating the students' briefs to identify large sections, footnotes, block quotes, headings, and legal citations.

2. In order to calculate the frequency of the attributes for the feature sets, splitting each brief-instance into sentences, breaking the sentences into word-tokens, and tagging those tokens for part of speech.

3. Calculating the frequency of the features selected and saving the resulting feature sets in a form usable by the machine learning platform I selected.

These processes are described in this appendix.

G.1 Manual annotation of texts

The first of these tasks involved manual annotation of the texts to identify segments of text that either would not be analyzed or those that would be held out from analysis at least temporarily. Working with a research assistant,[1] I developed a coding guide for manually annotating the papers in hard-copy using legal briefs other than those submitted by participants in this project. We then transferred our annotations from the paper copies to electronic copies using the General Architecture for Text Engineering or GATE (Cunningham et al., 2012). GATE is open-source software and available free of charge from the University of Sheffield.

Manual coding of paper copies involved two levels of the text's structure. At the large-segment level, we marked each of the following portions of each text, including any heading at the beginning of it:

- Caption: This is the formulaic block shown in Figure 4.1 at page 116.
- Tables: Though none of our student papers included tables of contents or tables of authorities, some courts require them in filed briefs and some attorneys provide them whether they are required or not. Because we developed the coding guide based on "published" attorney briefs, we had this segment type, but never used it in the context of this study.
- IntroSum: This included any introduction or summary immediately after the caption.
- Fact section, described above. See Section 4.3.1.
- Argument section, described above. See Section 4.3.1.
- Conclusion section, described above. See Section 4.3.1.
- OtherText: This is any material between the caption and the conclusion that does not fit any of the other large segments.
- OtherFormal: This is material before the caption or after the conclusion, usually consisting of pleading documents, such as motion and notice of motion, and the student's signature block after the conclusion.

[1] I am very grateful to the University of Minnesota College of Liberal Arts for a $5,000 Graduate Research Partnership Program grant in the summer of 2012 that made it possible for me to employ this research assistant.

See Appendix G.4 for the complete coding guide. I determined at once that I would not analyze materials (such as formal pleading documents) incidental to the brief and that I would not analyze the caption or signature sections because of their highly formulaic nature.

Within the large segments, we coded many other segments of text:

- Heading: The heading at the beginning of a section or subsection of the memo.
- Cite: This is any legal citation. These were coded depending on whether they were sentence citations (standing outside a textual sentence) or clause citations (appearing within a textual sentence). They were also coded by how many authorities were cited in a given citation. This measures (at least in part) the tendency of lawyers to employ "string cites," long citations of multiple authorities with little text to explain their purposes.
- BlockQuote: This is any quotation of 50 words or more, indented as required by legal writing conventions.
- Footnote: Any footnote reference or footnote text in the memorandum.

See Appendix G.4 for the complete coding guide. I decided to exclude section heads from my analysis, as I was uncertain of their linguistic status. I also excluded block quotations from my analysis, as they represent long stretches of text not composed by the students. I did not attempt to remove smaller quotations embedded within a student's text. So, for example, the following sentence appears in paper 1102:

> The general rule under the Copyright Act is that a "work protected under this title vests initially in the author or authors of the work."

My view is that such a sentence presents a hybrid of the student's language and the language of the quoted text because the student integrates her original composition with that of the quoted text. I did not attempt to identify the instances where students use such quotations frequently or where they constitute a large percentage of the student's paper.

We transferred this coding from the paper copies to electronic copies of the briefs in GATE. The process for doing so is described at length in the second half of Appendix G.4. I performed a check of inter-rater reliability on 10 of the papers (a little over 5% of them) to see whether the research assistant and I were consistently coding text spans the same way. I assessed inter-rater reliability using an F-measure with a β of 1. Using the IRR capability embedded in GATE, I assessed my codes as the *key set* and the research assistant's as the *response set*. Recall thus measured the percentage of spans that I annotated that were annotated in the research assistant's; precision, the percentage of spans the research assistant annotated that were annotated in my work. This effort is somewhat complicated by the fact that coders were not just assigning codes—which might be different depending on coder—to text spans, but they were also identifying the beginning and end of each span—which might not overlap exactly depending on coder. GATE provides for calculation of strict, lenient, or average agreement: For strict agreement, the text spans must overlap exactly and the codes assigned must be the same. For lenient agreement, if any part of the spans overlaps and the codes assigned are the same, the code is counted as a match. Average agreement counts codes where the spans do not overlap perfectly as half a match.

For this project, I was not worried about spans overlapping perfectly. If one coder included a space that the other did not, it was unlikely to affect the outcome of the project. On the other hand, it seemed very important that we were identifying the same spans on spans that would be excluded from analysis. When I did the analysis, I needed to be able to exclude citations and "OtherFormal" text from it completely, and the only way to do that is if we have carefully annotated them. So, I set these targets for IRR F-measures: strict $> .80$, lenient $> .95$, and average $> .90$. There is a variety of ways to run these tests with GATE, but our test instances met these thresholds in each case. When I examined the specific bases for disagreement, almost all were slight differences in span length, usually the inclusion or exclusion of a single space. I also noted that papers we annotated later had higher agreement than those annotated earlier.

G.2 Processing in Python and NLTK

I exported text and annotations from GATE in XML format. I wrote code in Python (Bird, Klein, & Loper, 2009) to merge the questionnaire data with the students' texts in GATE. Python is open-source software and is freely available on the Internet. See Appendix A for a list of files and resources that resulted from these steps available for download and use by other researchers. My Python code also excluded those segments of text described above that I did not wish to analyze, most notably headings, block quotations, and citations.[2]

Using the Natural Language Toolkit (NLTK), a freely available suite of natural language processing (NLP) tools written in Python (Bird et al., 2009), I wrote code to segment the instances into sentences, segment the sentences into words, and assign parts of speech to the words. I used the standard `sent_tokenize` module from the `nltk.tokenize` package. I used the `TreebankWordTokenizer` module from the `nltk.tokenize` package, which tagged tokens for part of speech using the Penn Treebank tagset (Atwell, n.d.). This last decision has implications for my ability to follow the Argamon/Koppel 02/03 study's approach and deserves further attention.

In the case of he Argamon/Koppel 02/03 study, the authors used the C5 tagset, because the texts they used were in the British National Corpus, which was already tagged for parts of speech with the C5 tagset. My corpus was not previously POS tagged, of course, so I needed to use a POS tagger. Most automated POS taggers use a combination of rule-based and statistical machine learning to assign tags to words in text (Jurafsky & Martin, 2009). Generally, this means they have to be trained on a corpus of text that is already tagged for parts of speech. The accuracy of the POS tagger is influenced by type of language used in the corpora on which it is trained. Generally speaking, a POS tagger applied to text in a specialized domain will perform better if it is trained on a corpus of texts from that domain. Thus, I would expect the best performance from a POS tagger applied to my corpus if it had been trained on a corpus of legal texts. However, I did not have access to such a corpus or previously trained POS tagger. I also wished to use the NLTK toolset, principally for my own convenience, as numerous natural language processing tools come ready-made in it. The `TreebankWordTokenizer` module uses the Penn Treebank tagset and is trained on the Penn Treebank, a large corpus of texts in American English on general topics.

As a consequence of my decision to use the NLTK `TreebankWordTokenizer` module, my corpus was not tagged by a POS tagger trained on legal texts in American English and it was not POS tagged with the same tagset as the Argamon/Koppel 02/03 study's.

[2] I actually generated three corpora, one consisting of the full texts of students' briefs, one consisting of just the fact sections, and one consisting of the full briefs with the fact sections deleted. This dissertation reports the analysis only of the first of these, as the other two provided very similar findings.

To address the former issue, I should perform an assessment of the accuracy of the NLTK `sent_tokenize` and `TreebankWordTokenizer` modules on my corpus, but I have not yet done so systematically. A brief, unsystematic review suggested that the POS tagger was working well.

The latter issue may raise some concerns about whether I have closely followed the Argamon/Koppel 02/03 study. The C5 tagset has 61 parts of speech, whereas the Treebank tagset has only 45. Some key differences are that the Penn Treebank includes a separate tag for each of a variety of punctuation marks, including the comma, period, right and left parenthesis, and dollar sign. It does not contain separate tags for the various forms of certain verbs (like *do, does, did*); nor does it make a distinction between prepositions and subordinating conjunctions. It is thus possible that an emphasis on more detail in certain areas in the C5 tagset and in other areas in the Treebank tagset could mean that my study is overlooking stylistic differences that were manifest in the Argamon/Koppel 02/03 study, or vice versa. I could address that question in the future by examining my corpus with the C5 tagset. For now, though, it's the Treebank tagset for this study.

I used Python to calculate the 100 most frequent POS bigrams and 500 most frequent trigrams in the corpus. On each text, my Python code then counted the number of occurrences of each of each function word, each POS tag, each of the 100 most common POS bigrams, and each of the 500 most common POS bigrams. Because the absolute frequency of each of these features in each text is likely influenced by the length of the text (a longer text has more nouns, etc.), I calculated relative frequencies in the same manner that the Argamon/Koppel 02/03 study did: For each function word and POS feature, I divided it by the number of tokens in the text instance, for each POS bigram by half the number of tokens in the text, and for each POS trigram by one quarter the number of tokens in the text.

G.3 Export of data to ARFF files for WEKA

Finally, I used Python code to write the corpus to an ARFF file, which was used by the WEKA machine learning framework. The resulting file represented 193 instances, each a text of one student, each classified by student's self-identified gender, and each described by 986 features.

G.4 Coding guides for manual annotation

The following pages provide the two coding guides for the manual annotation of texts in preparation to create the corpus for this study. The process is described briefly in Chapter 4 and in Appendix G.1.

There are two segments in this appendix. The first describes the process for identifying spans of text to code. The second describes the process of applying those coded spans to an electronic copy of each brief using the General Architecture for Text Engineering or GATE (Cunningham et al., 2012).

Paper coding guide

Researcher: Brian N. Larson
Revised July 9, 2013

Overview

The purpose of this coding project is to create annotated versions of memos written by law students and professional attorneys; the annotations identify parts of the memos like large sections, text headings, and citations to legal authorities. This annotation process will help to create a *corpus* (pl. *corpora*) of texts that the researcher will use for various projects.

The process will consist of two steps:

1. You will read and mark the memos in paper form.
2. You will record the annotations on a computer using software called "GATE: General Architecture for Text Engineering."

This document describes the first step, marking of memos in paper form.

Marking document segments

In this first phase, you will read and mark text segments in paper memos. For those segments that have types LargeSegment and Cite, you will indicate which type each instance is. For Cite segments, you will identify the number of authorities for each cite.

Use whatever hand annotations are convenient for you to mark the paper copies. You will sit down with the researcher and compare notes after you complete some samples.

You will find it helpful to mark each kind of segment separately. For example, mark beginning and ends of all LargeSegments before moving on to marking Cites.

Keep a separate journal about your experiences. Note any challenging coding in your journal, making reference to the paper number and page number when you have a problem. Assign codes in any case, using your best judgment and making a note. (It's better to assign codes even if you are unsure whether you should.)

Document segments defined

LargeSegment: A LargeSegment is a large 'chunk' of the text of the memo. Every memo is divided into several LargeSegments. Every portion of each memo is included in one LargeSegment or another. The following are the LargeSegements possible in these memos:

Caption: This is the 'top' of the memo as it would be filed with a court. It includes the name of the jurisdiction and court, usually in block capitals, the names of the parties to the litigation, and the title of the memo itself. This section is often (though not always) separated from the rest of the document by horizontal lines or a "box" around it. The title of the memo is included in this segment, even if it appears immediately following the horizontal line or box. In student memos, the Caption may be preceded by a title page that indicates a word count or other information; that front matter should be marked as LargeSegment:OtherFormal. This section appears in every memo.

TOCTOA: This type of segment includes any table of contents or table of authorities appearing in the document. A table of contents outlines the contents of the document, usually showing page numbers where headings appear. A table of authorities is a list of legal (and possibly other) authorities cited in the document.

IntroSum: This introduction or summary appears immediately after the Caption. It is usually two or three paragraphs at most. It usually has a header titled "Introduction," "Summary," and less commonly, "Procedural Background," but it may follow the caption directly without any header at all. It may include a sub-section titled "Issues," identifying the issues before the court. This section appears in almost every memo.

Facts: This section almost always follows the IntroSum section and is almost always titled "Facts" or "Factual Background," but may have names like "Stipulated Facts" or "Undisputed Facts." The key component here is the term "Fact" in the header and a recounting of the facts associated with the case. This section appears in every memo.

Argument: This section follows the Facts section. It usually, but not always, begins with a heading titled "Argument" or something similar. It may occasionally not be marked by such a heading but instead begin with a heading marking the beginning of the memo's argument. For example:

> **UNDER 17 U.S.C. § 101(2), THE WRITTEN INSTRUMENT NEED NOT BE SIGNED PRECEEDING CREATION OF THE WORK AS LONG AS THERE IS A PRIOR EXPRESS ORAL ARGEEMENT.**

This heading signals a shift from the fact section, which generally does not make reference to the law or to legal conclusions. If such a heading appears after the beginning of the Fact section, it signals the beginning of the Argument. This section appears in every memo.

Conclusion: This section appears after the Argument Section. (Some might consider it part of the Argument Section, but you should treat it as the beginning of a new section.) It usually begins with a heading titled "Conclusion" or words to that effect. Not all memos will have a separate conclusion section.

OtherFormal and OtherText: Any material that does not fit into the other LargeSegments identified here should be marked as "OtherFormal" or "OtherText."

OtherFormal covers any front matter before the Caption or after the Conclusion and other pages that may be styled as "Motion," "Notice of Motion and Motion," "Certificate of Service," "Proposed Order," or the like. It includes any signature block or formulaic closing where the attorney says "Respectfully submitted" (or words to that effect), indicates the date of the filing, and provides her name, firm, and contact information, where applicable. (Almost all memos will have such a signature block.)

OtherText is for segments of text between the caption and conclusion that do not appear to fit into other large segment categories, such as "Standard of Review," "Summary of legal principles," etc.

Heading: This indicates a heading at the beginning of a section or subsection of the memo. Headings are sometimes complete sentences. Headings are almost always set off from the surrounding text by one or more of the following typographical conventions: bold, Italics, underlining, centered, numbered or lettered sequentially. Headings may be marked by different typographical conventions within a single document; for example, one level of heading might be bold and another Italics. Do not identify headings within the Caption section. Headings may appear at different "levels," usually distinguished by the use of different typographical conventions or different numbering lettering. The heading is included in the LargeSegment it precedes. It is possible for to headings to appear consecutively, in which they should be coded as two consecutive headers; for example:

ARGUMENT

I. OGS PLED SUFFICIENT FACTS TO ESTABLISH A VALID CLAIM FOR COPYRIGHT INFRINGEMENT ON WHICH RELIEF CAN BE GRANTED

Cite: This is a reference to a text or authority outside of the memo. It may be a reference to a case, statute, or other authority. It is always set off from the grammatical portion of a sentence by commas or other punctuation. It is sometimes bracketed by parentheses. For example, the cites in the following sentences are highlighted:

> As children reach adolescence, courts recognize that the process of graining independence is an important consideration in determining duty and reasonable care. Restatement (Third) of Torts: Affirmative Duty § 42 (Tentative Draft No. 4, 2004).

> Lime is a well known screenwriter with fifteen years of experience in television writing, (Compl. ¶ 11.), and OGS sought to commission Lime to write an episode of *Lawless Love*, (*Id.* ¶ 8.).

The two preceding examples also illustrate two kinds of Cites:

Sentence: A Sentence Cite is punctuated as a complete sentence set off from the rest of the author's text. (Like the *Restatement* cite in the previous examples.) A sentence cite can be very short. (Id.¶24.) or (Id.). It is also possible for two citation sentences to appear in a row.

Clause: A Clause Cite appears within one of the author's prose sentences but it set off from it by commas (or sometimes a comma and a semi-colon). (Like the Compl. and Id. cites in the previous examples.)

A citation may include explanatory material in parentheses or an explanatory clause. Parenthetical and explanatory information is included in the cite and should be marked as part of the cite. Citations that appear in footnotes should be treated just like citations in the text.

Each citation refers to one or more authorities. When you mark a cite, you will also indicate the number of authorities identified in the cite (1, 2, 4, or 5 or more). When several authorities are identified in a cite, it is referred to as a "string cite," though you won't annotate it as such. See these examples (citations highlighted and number of sources identified in [brackets] at the end of the sample):

> In keeping with Plaintiff's contractual obligations, Vendor grants access to Plaintiff's content only to third parties that either subscribe to the Plaintiff Database or have obtained express written permission from a customer. (*See* Mem. Op. & Order 5; *cf. also* Countercl. ¶ 15 (alleging that "[i]nformation contained within databases is shared with other members"); *id.* ¶ 18 (alleging that vendors have the technical capability to grant access to data to third parties).) **[2 authorities]**

> Accepting this allegation as true for purposes of this motion, Plaintiff still does not overcome *Noerr-Pennington* immunity. "[I]t is clear that a defendant's invocation of adjudicative process to press legitimate claims is protected even though its purpose in doing so is to eliminate competition." *Razorback Ready Mix Concrete Co. v. Weaver*, 761 F.2d 484, 487 (8th Cir. 1985) (citing *Noerr*, 365 U.S. at 140; *Pennington*, 381 U.S. at 669); *see also MCI Commc'ns Corp. v. Am. Tel. & Tel. Co.*, 708 F.2d 1081, 1156 (7th Cir. 1983), *cert. denied*, 464 U.S. 891 (1983) ("Without a doubt, the intention to harm a competitor is *not* sufficient to make litigation or administrative proceedings a sham. That anticompetitive motive is the very matter protected under *Noerr-Pennington*."). [**2 authorities**, but note how the cited authorities also refer to authorities. So, the "cert. denied" is part of the citation to the *MCI* case because it describes the history of that case; similar explanatory marks include *aff'd, rev'd, overruled by,* etc.]

If the author of the memo refers to the name of an authority (or author) in an actual sentence of prose, that name is not a cite, but citation information following it within a sentence is a citation clause. In other words, if the name has a grammatical role in a sentence, it is not a citation. For example:

> This case differs from *Bjerke.* In *Bjerke,* the defendant provided a home away from the plaintiff's family, adopted many rules, and had extensive authority over the plaintiff's welfare. 742 N.W.2d at 665.

> In *Northwest Wholesale Stationers, Inc. v. Pacific Stationery & Printing Co.*, 472 U.S. 290 (1992), the Supreme Court described the essential attributes of a *per se* illegal boycott, including (1) joint efforts by a firm or firms to disadvantage competitors, and (2) the conspirators' possession of "market power or exclusive access to an element essential to effective competition." *See id.* 294-96.

In this example, the references to "*Bjerke"* and "*Northwest Wholesale…"* are not Cites, but the citation sentences at the ends of the sentences are, as is the clause after the first instance of the name of the *Northwest Wholesale* case.

BlockQuote: When a memo includes a long quote from another authority, the author sets it off from the surrounding text by indenting it on the left and perhaps by single-spacing it. (Most of the memo text will be double-spaced.) Do not annotate citations in block quotes.

Footnote: The author may have place a footnote reference in the text of her memo and display a footnote at the bottom of the page. The footnote reference number in the text and the reference number and footnote text at the bottom of the page should both be marked as "Footnote." Any citations in a footnote should be annotated as citations.

GATE annotation guide

Researcher: Brian N. Larson (BNL)
Revised July 14, 2013

Overview

The purpose of this coding project is to create annotated versions of memos written by law students and professional attorneys; the annotations identify parts of the memos like large sections, text headings, and citations to legal authorities. This annotation process will help to create a *corpus* (pl. *corpora*) of texts that the researcher will use for various projects.

The process will consist of two steps:

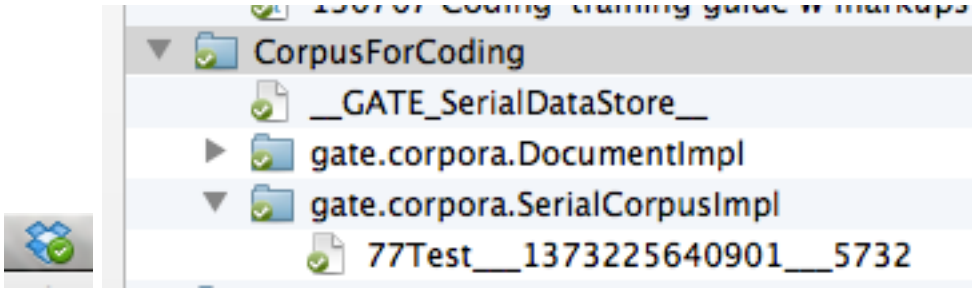
1. You will read and mark the memos in paper form.
2. You will record the annotations on a computer using software called "GATE: General Architecture for Text Engineering."

This document describes the second step, putting the annotations on the document in GATE. Follow these instructions whenever you are annotating documents for BNL using GATE. ***These instructions are designed with the assumption that you'll have at least half an hour to work or so; that's because the instructions for initiating an closing a coding session are a little cumbersome, and if you do many short coding sessions, you'll end up spending too much of your time starting up and shutting down.***

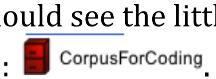
Beginning a session

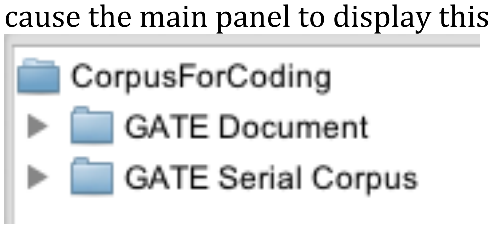
In each of your work sessions, follow this process:

1. Make sure that BNL is not coding at the same time. Do this by checking your email and seeing whether he as sent you an email saying he is coding or an email indicating that he has finished.
2. Send an email to BNL indicating that you are beginning to code (so he knows not to).
3. Make sure that your Dropbox sync is completed (the files you work on are local on your computer, but they must be sync'ed to Dropbox to ensure you have the most recent copy). Look for green check at top of screen and green checks on the relevant folders:



4. Open GATE.
5. On the “Messages” screen at startup, be sure that the plugins “ANNIE” and “Scheme_Annotation_Editor” are loaded.
6. Load the five annotation schemas. For each, go to “Language Resources,” right click, choose “New,” and then “Annotation Schema.” The schemas are in the folder titled “GATE materials.” Load each of the following files with the names given.
    a. LargeSegmentSchema.xml, “LargeSegment”
    b. HeadingSchema.xml, “Heading”
    c. FootnoteSchema.xml, “Footnote”
    d. CitationSchema.xml, “Citation”
    e. BlockQuoteSchema.xml, “BlockQuote”
7. Right click on “Datastores” and choose “Open Datastore.” Follow these instructions:
    a. Choose “SerialDataStore…”
    b. Navigate to the Gender-Genre Team Annotation folder, click on the “CorpusForCoding” folder once, and choose “Choose.”
    c. Click on the little triangle to the right of “Datastores” 

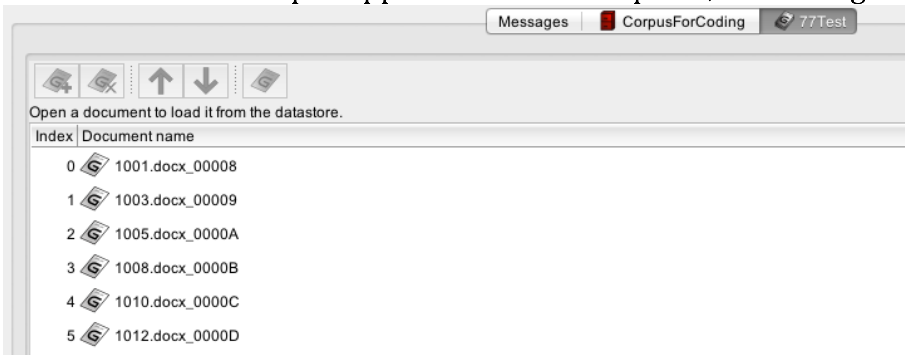
if necessary so that it points downward. (This is called “expanding” Datastores.) You should see the little file cabinet labeled “CorpusForCoding”: 

. Double click on it, which should cause the main panel to display this:


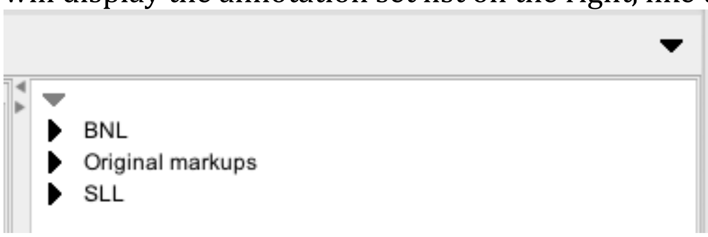
    d. Expand “GATE Serial Corpus” so that it displays “77Test.” Double-click on that.
    e. “77Test” should now show up in the Language Resources.
    f. Double-click on the “77Test” appearing under in the Language Resources. That should reveal a list of documents in the main panel, numbered according to the file numbers, with a code added. Thus, a file called “1001.docx” might appear as “1001.docx_0008.”
    g. You are ready to work on coding a document.

Coding a document

1. Click or double click on "77Test" under Language Resources so that the list of documents in the corpus appears on the main panel, something like this:

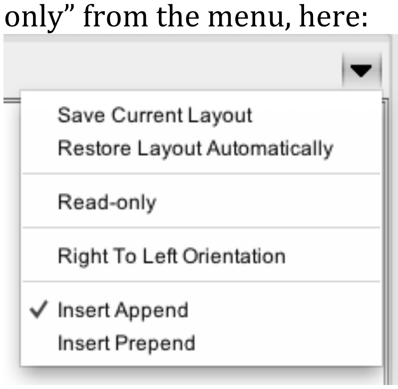


2. Double click on the document you want to edit. That will open it in GATE.
3. In the upper left-hand of the main panel, click on "Annotation Sets," which will display the annotation set list on the right, like this.



In this example, there are three markup sets. You may find that there is only one ("Original markups"). If there is not an SLL markup set, create one by typing "SLL" at the bottom of this pane and hitting "New."

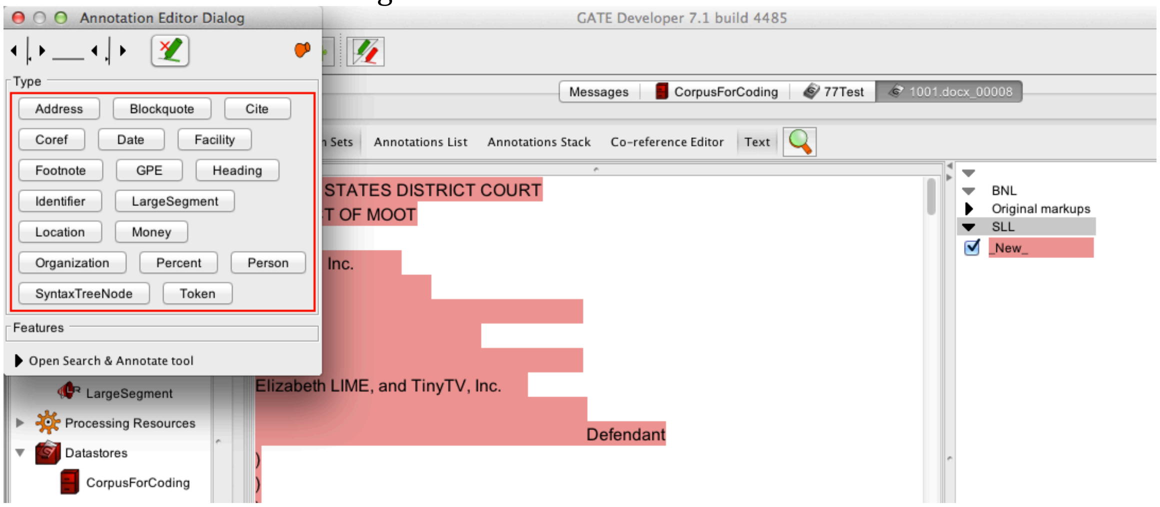
4. Using the "down arrow" in the upper right of the main display, select "Read only" from the menu, here:



This prevents you editing the underlying text, but permits you annotate it as required here.

5. It's probably easiest to annotate all segments of a given kind (LargeSegment, Heading, Cite, Footnote, etc.) at once, since GATE assumes that you want a new annotation to be the same as the last you gave.
6. Before adding any annotation, be sure that the SLL annotation set is selected on the right (that prevents you entering annotations as if they are mine).

7. Select the first segment you want to annotate, and then hover the mouse over it. You should see something like this:

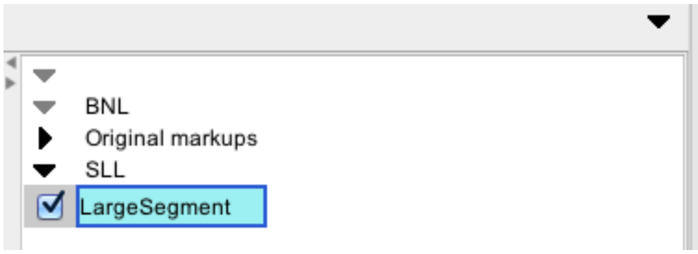


You must chose an annotation type from the Editor dialog. Some of these are defaults of GATE; don't use them. Use only the ones defined in the coding guide. I would start with the LargeSegment type to get it out of the way.

8. When you select an annotation type, you will generally have to fill out one or more "Features." Here are the possible features for LargeSegment:

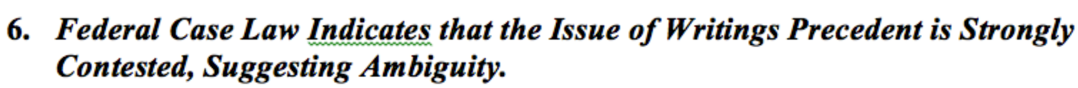
The red box around the feature means that it is required that you choose one.

9. Once there are no red boxes remaining in the Annotation Editor Dialog, close it by clicking the red circle in the upper left of it.
10. WARNING: Sometimes the Annotation Editor Dialog (AED) pops up before you make a new selection or finish making your selection. That's because whenever you "mouse over" an existing annotation that is visible on screen (i.e., it is colored), the AED opens so you can "inspect" the annotation. If that happens, be sure to close the AED and make sure you are hovering over where your new annotation will be when the AED pops up. (It's easy to accidentally replace an annotation you've already made.)
11. When you've added annotations of a particular type, that type shows up with a color code and check box under your annotation set:

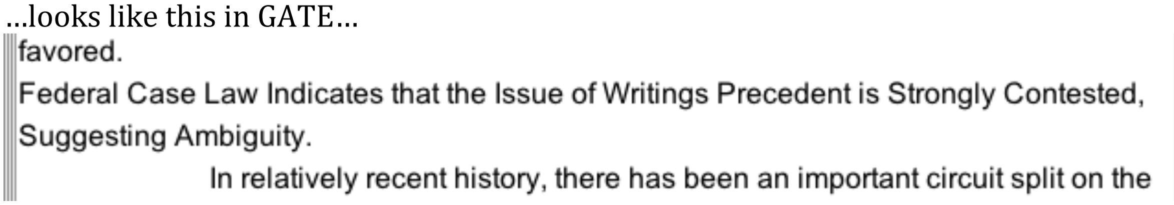


Hide annotations you are not currently using by unchecking their boxes. TIP: If you finish adding your LargeSegments, then click on this box, and then try to add your first Heading, GATE defaults the new annotation to LargeSegment, which causes the LargeSegment check to reappear in the box and makes the LargeSegment codes visible. You can just click on it again, and the next time, GATE will default to Heading, or whatever was the next code you added.

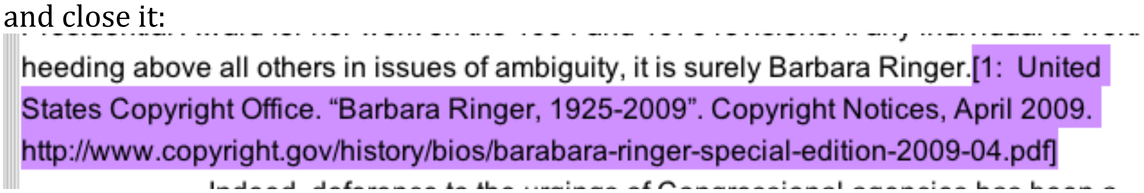
12. Add further annotations.
13. TIP: If you have trouble finding something you need to annotate, click the magnifying glass and search for a word in or near the span of text you are looking for. (Unusual words work better, of course!)
14. When annotating headings, note that GATE will not import the numbers on the automatically numbered heading paragraphs from Word. What looks like this in Word…



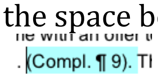
…looks like this in GATE…



15. When annotating a footnote, be sure to grab the square brackets that open and close it:



16. When doing citation annotations, follow these conventions for covering the surrounding punctuation:
    a. For well-formed sentence citations, start your span with the first number, letter or parenthesis of the citation and end your span with the space before the next sentence begins:

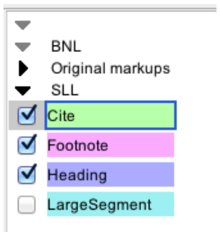
    

    b. For well-formed clause citations, start your span with the comma that begins the citation. Include the comma that ends it only if that comma is not necessary for the sentence to be properly punctuated absent the citation. Do not include a sentence-ending period in your citation span. Examples:

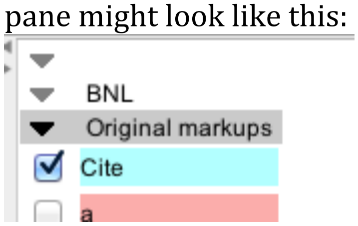
> Indeed, deference to the urgings of Congressional agencies has been a long-standing tradition recently reaffirmed by the Supreme Court in U.S. v. Mead Corp ., 533 U.S. 218 (2001). Quoting from the historic Skidmore v. Swift & Co ., 323 U.S. 134 (1944), the court iterated that "an agency's interpretation may merit some deference whatever its form, given the "specialized experience and broader investigations and information" available to the agency, id . at 139, and given the value of uniformity in its administrative and judicial understandings of what a national law requires, id., at 140 . "

c. Generally, with citations, whether well formed or ill-formed, sentence or clause, try to select the citation span so that what would be left if the citation were deleted would be a grammatical sentence, properly punctuated. Examples:
<<still coming>>

17. When you are done doing your annotations on this document, you should find that you have a set of annotation types and their colors displaying in the annotation sets pane under SLL, like this:



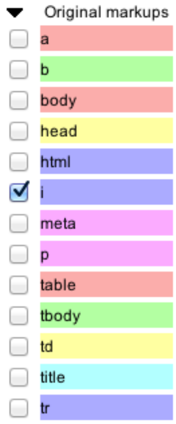
Unless there were annotations under "BNL" before you started annotating this document, there should be none now. All your annotations should appear under the SLL set. Unfortunately, if they appear elsewhere, the only way to get rid of them is to right click on the offending entry and delete it (along with all annotations associated with it). Say you accidentally put one of your citation annotations under Original markups. Your annotation sets pane might look like this:



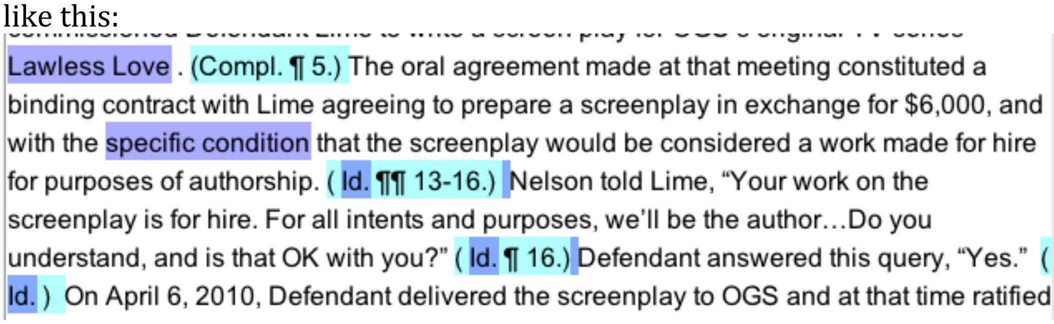
Right-click on the "Cite" label here and choose Delete. You'll need to go back through the document and reapply any citation annotations you have just deleted.

18. Take a few moments to browse back over your annotations to see that they look correct.
19. TIP: You can see whether you have missed italicized "Id." anywhere in the document by showing the original markup for Italics and underlining.

Expand “Original markups” in the Annotation sets panel, then click on “i” and “u” if they are visible, like this:



Then when you look through the document, you can see that your annotations of “Id.” citations will generally overlap an “i” or “u” annotation, like this:


Lawless Love . (Compl. ¶ 5.) The oral agreement made at that meeting constituted a binding contract with Lime agreeing to prepare a screenplay in exchange for $6,000, and with the specific condition that the screenplay would be considered a work made for hire for purposes of authorship. (Id. ¶¶ 13-16.) Nelson told Lime, “Your work on the screenplay is for hire. For all intents and purposes, we’ll be the author…Do you understand, and is that OK with you?” (Id. ¶ 16.) Defendant answered this query, “Yes.” (Id.) On April 6, 2010, Defendant delivered the screenplay to OGS and at that time ratified

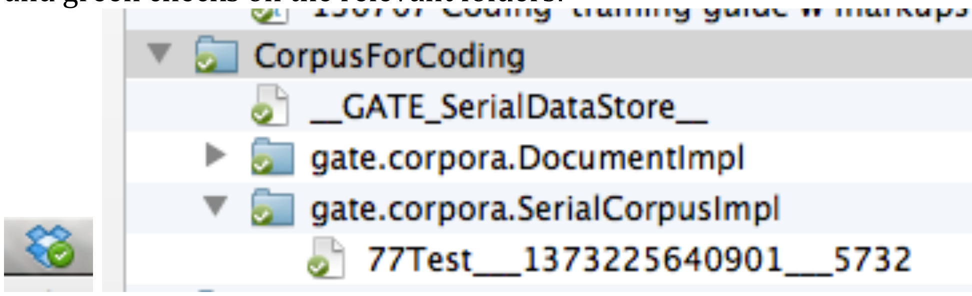
20. **REALLY IMPORTANT BIT:** When you have finished with this document, go to its name in the left pane in Language Resources, right-click and choose “Save to its Datastore.” Nothing is saved properly until you do this!
21. After saving, go to the document’s name in the left pane in Language Resources, right-click and choose “Close.”
22. On the paper copy of the document, note the date and time you finished putting annotations on the computer. (We made need this for disaster recovery. If we lose a day’s work, we need to know which documents we worked on that day.) Retain the paper copy—give it to BNL at your convenience.
23. Repeat this process with the next document.

For an example of a document where BNL has already layered annotations into it in GATE, open paper 1001, 1003, 1005, 1008, or 1010 and view the BNL annotation set.

Ending a session

1. Save the last document you worked on and close it.

2. In the left pane in Language Resources, right-click "77Test" and choose "Save to its Datastore."
3. Then right-click "77Test" and choose "Close."
4. Right-click "CorpusForCoding" under Datastores and close it, too.
5. Exit GATE; make sure the program has completely quit (it no longer appears in the list when you Command-TAB).
6. Wait for your Dropbox sync to be completed; green check at top of screen and green checks on the relevant folders:



7. Send an email to BNL indicating that you are done coding (so he knows that he can work on the corpus).

Appendix H

Examples of bigram and trigram features in context

This appendix provides examples of the bigram and trigram features discussed in Chapter 5 in the context of sentences in the corpus in which they appeared. The features are arranged alphabetically by feature label. The feature labels are made up of part-of-speech tags. See Appendix B for explanations of the POS tags. The feature is italicized within the context of the example sentence. The number following each example sentence indicates which paper it comes from (before the colon) and which sentence it is in that paper (after the colon). Mindful of the length of this dissertation, I have generally chosen shorter sentences for these examples where possible. In some cases, I have provided multiple examples, and in others, I have provided examples where it appears that the automatic part-of-speech tagging used in this dissertation resulted in questionable or incorrect POS tags.

1. **CC—DT—NN** (POS trigram, 1,351 instances in corpus)
   The legislative history of the statute confirms this interpretation *and this reading* advances the policy concerns of Congress. 1063:52
2. **CD—CD** (POS bigram, 3,295 instances in corpus)
   OGS and Lime entered a valid work for hire agreement under SS *101(* 2) of the Copyright Act. 1066:146 (Note: The POS-tagger counted the "101" and the "(" here as one CD—CD bigram, and then the "(" and "2" as a second, etc. This appears to be an error and explains the very large number of this bigram. "SS" represents a section mark.)

3. **CD—Comma—NNP** (POS trigram, 567 instances in corpus)
   On March 27, *2010, Dave* Nelson, executive producer of Lawless Love , met with Lime in the office of OGS. 1070:5
4. **CD—DT—NN** (POS trigram, 244 instances in corpus)
   Under the Copyright Act of 1976, a work-made-for-hire is either 1*) a work* prepared by an employee within the scope of his employment, or 2*) a work* specially ordered or commissioned for use in one of nine narrowly enumerated categories accompanied by a written agreement that the product will be considered a work made for hire. 1080:47
5. **Colon—DT—NN** (POS trigram, 303 instances in corpus)
   After some debate, the language remained*; no party* disagreed with her interpretation of the language. 1086:55 (Note: Semi-colons are correctly counted as colons by the POS tagger.)
6. **Comma—CC** (POS bigram, 3,511 instances in corpus)
   Furthermore, one of the main themes of the Copyright Act as a whole*, and* this provision in particular, is that of clarity and certainty. 1086:89
7. **Comma—DT—NN** (POS trigram, 4,261 instances in corpus)
   Plaintiff, creator of the show Lawless Love, contacted Elizabeth Lime*, a television* writer, on March 27, 2010, to discuss the possibility of her writing an episode for the show. 1087:8
8. **Comma—DT—NNP** (POS trigram, 1,130 instances in corpus)
   In September 2010, Lime and TinyTV broadcasted a pilot episode of a new television series*, The Tin* Can. 1087:18
9. **Comma—DT—NNS** (POS trigram, 433 instances in corpus)
   In order to survive a Motion to Dismiss*, the facts* as pleaded must state a plausible claim for relief, meaning they must assert more than a possibility that the defendant is liable. 1087:24
10. **Comma—PRP—MD** (POS trigram, 370 instances in corpus)
    For all intents and purposes*, we'll* be the author. 1088:13 (Note: The contraction "we'll" counts as two tokens.
11. **DT—JJ—IN** (POS trigram, 327 instances in corpus)
    It is common and reasonable for parties to take *each other at* their words, especially when a writing memorializing their agreement takes place a week after their oral agreement. 1091:27
12. **DT—JJ—NNS** (POS trigram, 956 instances in corpus)
    Plaintiff has succeeded in alleging *the necessary facts* to state a plausible claim upon which relief can be granted. 1091:32
13. **DT—NN** (POS bigram, 51,277 instances in corpus)
    Interpreting statutes is a matter of determining legislative intent, and where *the language* of *the statute* makes "*the intent*... plain, nothing is left to construction". 1091:37
14. **DT—NN—CC** (POS trigram, 1,949 instances in corpus)
    OGS failed to meet the requirements of *the statute and* therefore the copyright remains with the author. 1092:51

15. **DT—NN—Colon** (POS trigram, 323 instances in corpus)
Plaintiff may argue that *either party*—the original author or the hiring party—could be bringing a case for infringement. 1095:151 (Note: The POS tagger correctly tagged the '—' as a colon.)

16. **DT—NN—Comma** (POS trigram, 3,359 instances in corpus)
If the original author publicized the work before the execution of *the writing,* intending to sign something at a later date, has he or she willfully infringed a copyright? 1095:157

17. **DT—NN—DT** (POS trigram, 333 instances in corpus)
The Second Circuit held that both parties must "underst[an]d at *the time the* works were created that the works were made for hire. 1096:134

18. **DT—NN—IN** (POS trigram, 16,075 instances in corpus)
The court found that text printed on *the back of* checks stating that Playboy owned "right, title, and interest" in the item "was insufficient for the SS 101(2) writings requirement. 1096:136

19. **DT—NN—JJ** (POS trigram, 8,560 instances in corpus)
Nelson's oral statement was ambiguous, and Lime may not have understood his intent to enter into *a work-for-hire agreement.* 1096:142

20. **DT—NN—MD** (POS trigram, 2,064 instances in corpus)
The commissioning party gains the security of knowing *the artist will* not take the sell copies of the work to a competitor. 1096:153

21. **DT—NN—NN** (POS trigram, 6,749 instances in corpus)
And third parties can more easily ascertain *the copyright status* of works they wish to buy or use. 1096:155

22. **DT—NN—NNS** (POS trigram, 814 instances in corpus)
Here, the Copyright Act provides for *both civil remedies* and criminal penalties. 1096:166

23. **DT—NN—OQuote** (POS trigram, 436 instances in corpus)
On April 2, 2010, OGS sought copyright protection for *the screenplay "*Anticipatory Repudiation" authored by Ms. Lime. 1110:10

24. **DT—NN—Period** (POS trigram, 3,928 instances in corpus)
SS 101(2) requires a written agreement signed prior to the creation of *the work.* 1110:18

25. **DT—NN—POS** (POS trigram, 1,143 instances in corpus)
"Shall" is defined as "expressing *the speaker's* determination to bring about... 1110:44

26. **DT—NN—PRP** (POS trigram, 286 instances in corpus)
In exchange for $6,000, Ms. Lime signed the agreement which stated *the screenplay she* had prepared was a work made for hire and OGS was the author of the screenplay. 1111:14

27. **DT—NN—RB** (POS trigram, 809 instances in corpus)
The introductory statement, "*a work specially* ordered or commissioned," signifies the creation is to take place in the future. 1111:42

28. **DT—NN—TO** (POS trigram, 2,147 instances in corpus)
The implied sense of future tense in the language of the statutory definition combined with the use of “shall,” indicates the drafters meant for *a writing to* be signed prior to creation. 1111:45

29. **DT—NN—VBD** (POS trigram, 3,239 instances in corpus)
The legislature intended that work for hire agreements be signed prior to creation in order to protect the rights of the creator. 1111:102

30. **DT—NN—VBG** (POS trigram, 676 instances in corpus)
Defendants’ move for Judgment on the Pleadings because there was no express written agreement signed by *either party preceding* the creation of the screenplay. 1114:14

31. **DT—NN—VBN** (POS trigram, 1,894 instances in corpus)
OGS requests relief in the form of a preliminary injunction against TinyTV and Lime from showing *a production based* on the copyrighted screenplay belonging to OGS. 1115:4

32. **DT—NN—VBZ** (POS trigram, 3,534 instances in corpus)
The language of *the statute is* not ambiguous enough to support such a narrow reading. 1115:118

33. **DT—NN—WDT** (POS trigram, 534 instances in corpus)
The Plaintiff explicitly commissioned the Defendant Lime to create a screenplay that the Plaintiff would own all rights to. 1116:159

34. **DT—NNP** (POS bigram, 8,966 instances in corpus)
The Plaintiff explicitly commissioned *the Defendant* Lime to create a screenplay that *the Plaintiff* would own all rights to. 1116:159

35. **DT—NNP—CC** (POS trigram, 261 instances in corpus)
The interpretation of *the First and* Second Circuit should be preferred over that of the other circuits. 1118:156
Those conditions were for *both Nelson and* Lime to sign a written instrument expressly stating that the work was created as a work–made–for–hire. 1120:68

36. **DT—NNP—Comma** (POS trigram, 310 instances in corpus)
As the ownership of the copyright for the screenplay ”Anticipatory Repudiation” was not expressly transferred to *the Plaintiff,* Defendant Lime retains ownership of the copyright. 1120:78

37. **DT—NNP—IN** (POS trigram, 727 instances in corpus)
Even when considering the allegations made in *the Complaint in* a light most favorable to the Plaintiff, the Court should find that there no plausible entitlement for relief. 1120:130 (Note the practice of legal writers to capitalize the names of pleadings in the case. The POS tagger treated this as a proper noun.)

38. **DT—NNP—MD** (POS trigram, 767 instances in corpus)
Therefore, *the Court should* grant Defendants’ motion to dismiss. 1120:131 (Note the practice of referring to the court to which the brief is directed with a capitalized noun. The POS tagger treated this as a proper noun.)

39. **DT—NNP—NN** (POS trigram, 789 instances in corpus)
    Simmons spoke to the defendant's prior to 2010 about including *the AGA safety* manual in the trampolines they sell in response to two injuries that occurred on defendant's trampolines. 2004:66

40. **DT—NNP—NNP** (POS trigram, 3,334 instances in corpus)
    Following to providing the defendants with *the American Gymnastic* Association's manual, Simmons was not contacted by them again. 2004:67

41. **DT—NNP—Period** (POS trigram, 381 instances in corpus)
    Therefore, Lawzz.com respectfully requests that this court grant summary judgment in favor of *the Defendant.* 2005:103

42. **DT—NNP—POS** (POS trigram, 523 instances in corpus)
    Therefore, Mr. Sullivan respectfully asks this court to dismiss *the Defendant's* motion for summary judgment. 2006:48

43. **DT—NNP—VBD** (POS trigram, 735 instances in corpus)
    After a short investigation, *the Defendant discovered* that Brandon McAllister was the creator of the game, and by the end of the day he had met with Brandon, who admitted to being the creator. 2007:17

44. **DT—NNP—VBZ** (POS trigram, 633 instances in corpus)
    Defendant was warning Plaintiff against the danger outside *the OCC premises.* 2009:109

45. **DT—NNS** (POS bigram, 7,455 instances in corpus)
    Defendant used the word warning to show the seriousness of the danger outside that Plaintiff would be likely to encounter if he left *the premises.* 2009:113

46. **DT—NNS—CC** (POS trigram, 497 instances in corpus)
    Lawzz.com reasonably communicated its terms of use; therefore, Plaintiff had constructive notice of *those terms and* entered into an enforceable contract. 2011:56

47. **DT—NNS—Comma** (POS trigram, 559 instances in corpus)
    In *both cases,* the consumer is prompted to examine terms of sale that are located somewhere else. 2011:76

48. **DT—NNS—IN** (POS trigram, 2,069 instances in corpus)
    Here, *the terms of* use are displayed on every page, simultaneously with website use, and once again before the first time a user stores data. 2011:83

49. **DT—NNS—Period** (POS trigram, 537 instances in corpus)
    Plaintiff was a frequent user and had constructive notice of *the terms.* 2011:96

50. **DT—NNS—POS** (POS trigram, 249 instances in corpus)
    The neighbors' complaints of The Clubs actions lead to The Stein to prohibit The Club from gathering at their establishment before and after rides, not Sweeney. 2018:136

51. **DT—NNS—RB** (POS trigram, 333 instances in corpus)

    Defendant has denied *all allegations specifically* stating that the parties did enter into a valid contract. 2022:31

52. **DT—NNS—TO** (POS trigram, 304 instances in corpus)

    Because there is nothing in *the facts to* support this conclusion, more discovery is necessary to form a valid conclusion for this element. 2022:110

53. **DT—NNS—VBD** (POS trigram, 994 instances in corpus)

    For *the reasons set* forth above, Defendant's motion for summary judgment must be granted. 2023:154

54. **DT—NNS—VBP** (POS trigram, 662 instances in corpus)

    As Plaintiffs were leaving the restaurant, Mr. Mohammed heard one of *the customers say*, "good, *the terrorists are* leaving. 2026:31

55. **IN—DT—NN** (POS trigram, 22,798 instances in corpus)

    Defendant made no remark *at that time.* 2026:33

56. **IN—DT—NNP** (POS trigram, 3,899 instances in corpus)

    Bill's Bar and Grill meets the definition of a public accommodation *under the Minnesota* Human Rights Act. 2026:55

57. **IN—DT—NNS** (POS trigram, 3,688 instances in corpus)

    An analysis of each *of these elements* will demonstrate that Plaintiffs have not established a genuine issue of material fact. 2026:74

58. **IN—NNP—NN** (POS trigram, 271 instances in corpus)

    Since Plaintiffs fail to prove discriminatory motive, they are unable to succeed on this element. 2026:172 (Note: This appears to be an error by the POS tagger. Here, "fail" should be tagged as a present-tense verb.)

    As such, Harting is at odds *with Minnesota law* pertaining to primary assumption of risk. 2030:128

59. **IN—PRP$—NNS** (POS trigram, 348 instances in corpus)

    After assembling the trampoline, Sullivan read the trampoline manual aloud *with his children* and wife. 2033:46

60. **JJ—DT—NN** (POS trigram, 507 instances in corpus)

    Plaintiff only thinks now that it could not be possible that he would have answered Defendant's questions differently than when he was *drunk the night* in question. 2034:48

61. **JJ—NN** (POS bigram, 16,322 instances in corpus)

    Defendant prays the Court to take into account that Plaintiff has not offered *substantial evidence* that there exists a question of fact for a jury. 2034:59

62. **JJ—NNS** (POS bigram, 4,820 instances in corpus)

    Thus, Defendant is entitled to official immunity due to her *discretionary actions* being the actual subject of suit. 2034:69

63. **JJ—TO—VB** (POS trigram, 831 instances in corpus)
If official immunity is going to function as intended and protect officials' decisions from judicial second-guessing, then officials must not be *hesitant to make* good faith decisions that are sometimes, in hindsight, based on imperfect knowledge. 2034:103

64. **NN—Comma** (POS bigram, 10,557 instances in corpus)
A facial *expression,* by itself, is not enough to prove malice. 2034:111

65. **NN—Comma—NNP** (POS trigram, 1,244 instances in corpus)
While waiting for the *chairlift, Plaintiff* was struck by John Francis, who was skiing near Plaintiff. 2035:9

66. **NN—DT—NN** (POS trigram, 1,699 instances in corpus)
The court stated that, "[w]hether a duty exists depends on the relationship among parties and the foreseeability of harm to others. 2035:67 (Note that the POS tagger did not recognized the amended quotation and thus tagged "hether" as a noun.)
At the care *center a bandage* was put over Mr. Palumbo's eye and an ambulance was called to take Mr. Palumbo to the hospital. 2036:38

67. **NN—IN—NNP** (POS trigram, 3,531 instances in corpus)
On November 3, 2011, Plaintiff filed the Complaint praying for *judgment against Defendant* in an amount in excess of $50,000. 2036:40

68. **NN—NN—CC** (POS trigram, 766 instances in corpus)
Because Ms. Daniels had not heard back from *law enforcement nor* had the guests been informed of the situation, she requested that no one leave until OCC had received an "all clear" from police. 2037:25

69. **NN—NN—MD** (POS trigram, 441 instances in corpus)
While the incident that occurred in the *school cafeteria may* be traceable to Charles's haircut, the haircut cannot be considered a substantial or material disruption under the standard established in Tinker. 2040:162

70. **NN—TO** (POS bigram, 6,099 instances in corpus)
The School District produced little *evidence to* explain the nexus between the haircut and the incident. 2040:163

71. **NN—TO—VB** (POS trigram, 3,369 instances in corpus)
The School District produced little *evidence to explain* the nexus between the haircut and the incident. 2040:163

72. **NN—VBZ** (POS bigram, 6,455 instances in corpus)
Therefore, as a matter of law, the school *district is* not entitled to judgment and the school district's motion for summary judgment should be denied. 2040:182

73. **NN—VBZ—IN** (POS trigram, 866 instances in corpus)
The phrase used for Charles's *haircut is from* a national organization. 2040:184

74. **NNP—CC** (POS bigram, 2,241 instances in corpus)
Ruby Jones previously traveled on a World Transformational Tours tour of *Ethiopia and* Somaliland without incident and heard that the agency offered a unique tour focused on the politics and society of Namibia. 2042:3

75. **NNP—CD** (POS bigram, 2,380 instances in corpus)
Ruby Jones paid the remaining balance of 10,000 dollars before the *August 30*, 2011 deadline. 2042:15

76. **NNP—Comma—NNP** (POS trigram, 1,091 instances in corpus)
While on the planned seven-hour layover at Lagos International Airport in *Nigeria, Ruby* Jones and thirty-six other passengers were taken hostage by the Nigerian Resistance Army. 2042:20

77. **NNP—MD—VB** (POS trigram, 1,272 instances in corpus)
Next, there are genuine issues of material fact as to whether Plaintiff has met any of the three elements *Defendant must show* in order to invoke the primary assumption of the risk doctrine. 2043:7

78. **NNP—NNP—Comma** (POS trigram, 1,583 instances in corpus)
The incident occurred at approximately 3:18 p.m., shortly after Plaintiff and his companion, *Jennifer Gray,* returned to skiing after taking a break for lunch. 2043:12

79. **NNP—NNP—IN** (POS trigram, 1,562 instances in corpus)
However, the *Supreme Court of* Minnesota has since reaffirmed Schroeder in 2008 and 2010 decisions. 2044:49

80. **NNP—TO—VB** (POS trigram, 520 instances in corpus)
These secondary disruptions related to the incident were enough for the *Court to find* that the ban on Confederate flag images in Barr was warranted. 2044:80

81. **NNP—VBD—TO** (POS trigram, 579 instances in corpus)
When Mr. *Swensen got to* the bottom step, he hooked his foot, fell and injured his left knee and leg. 2046:25

82. **NNP—VBZ—IN** (POS trigram, 462 instances in corpus)
First, Mr. *Johnson asserts that* the Defendant's lighting of the stairway was inadequate which contributed to the fall. 2046:29

83. **NNS—Comma—IN** (POS trigram, 269 instances in corpus)
As Castorina *highlights, in* order for a conflict to make disruption reasonably foreseeable there must be indisputable proof of racially-driven incidences. 2047:116 (Note that the POS tagger has erroneously tagged the verb "highlights" as a noun.)
Ministerial *duties, as* opposed to discretionary duties, are those which are "absolute, certain and imperative, involving merely the execution of a specific duty arising from fixed and designated facts. 2051:48

84. **NNS—DT—NN** (POS trigram, 645 instances in corpus)
Like Gernander, using the words "I think" *constitutes an opinion.* 2053:84 (Note that the POS

tagger has erroneously tagged the verb “constitutes” as a noun.)

This change was made pursuant to numerous calls and *complaints the tavern* had received from neighbors, as well as Sweeney, regarding the excessive noise created by members of the Club. 2054:37

85. **NNS—MD—VB** (POS trigram, 740 instances in corpus)

    Although the *statements may be* construed as uncomplimentary, they do not suggest verifiably false facts about the Plaintiff. 2054:66

86. **NNS—TO—VB** (POS trigram, 831 instances in corpus)

    Failure to recognize such a privilege would place a burden on public *officials to weigh* the necessity of efficiently addressing local disturbances and/or concerns against the potential civil liability for doing so. 2054:127

87. **OQuote—DT—NN** (POS trigram, 890 instances in corpus)

    A summary judgment may be granted when *“either party* is entitled to a judgment as a matter of law. 2056:55

88. **OQuote—DT—NNP** (POS trigram, 409 instances in corpus)

    Two corporations entered into a lease agreement (*“the Lease*”) in which American Golf Inc. (“AGI”) was asked to provide a service to Golfman Institute Inc. (“Golfman”). 2066:0

89. **POS—NN—TO** (POS trigram, 403 instances in corpus)

    Additionally, the doctrine of loco parentis is far from absolute in a school*’s ability to* punish a child in place of a parent. 2070:105

90. **PRP—VBZ—DT** (POS trigram, 285 instances in corpus)

    In reality, *it is an* expression of the Defendant’s realization that it cannot ensure the safety of its clients against all possible risks they could encounter while travelling. 2071:4

91. **RB—DT—NN** (POS trigram, 1,356 instances in corpus)

    These acts are *precisely the type* of accident that WTT’s disclaimer and release disclaims liability from. 2071:71

92. **RB—IN** (POS bigram, 2,501 instances in corpus)

    Like the tour operator in Powell, WTT was *not in* a joint enterprise with the airport where Plaintiff was injured. 2071:73 (There were 112 instances of “not in” in the corpus.)

    In fact there is no mention of intentional, willful, or wanton conduct *anywhere in* WTT’s disclaimer and release. 2071:79

93. **RB—IN—NN** (POS trigram, 300 instances in corpus)

    The disclaimer and release provided by WTT to Plaintiff, prior to the Namibia Politics and Society Tour 2011, was *not in violation* of public policy. 2071:111

94. **RB—TO—VB** (POS trigram, 432 instances in corpus)

    Atticus informed M. Atticus that Wager expressed J. Atticus being “old *enough to take* care of himself.” 2072:47 (Note that this trigram appears inside quoted text.)

95. **TO—DT—NN** (POS trigram, 2,541 instances in corpus)
J. Atticus never exhibited violence toward anyone prior *to the stabbing.* 2072:56

96. **TO—DT—NNP** (POS trigram, 299 instances in corpus)
In Tinker, three students wore black armbands to school as a symbolic expression publicizing their objections *to the Vietnam* War. 2074:69

97. **TO—DT—NNS** (POS trigram, 321 instances in corpus)
Similar *to the students* in Tinker who wore the armbands to promote awareness of their views and to encourage others to adopt them, Charlie was just trying to support his sister and raise awareness about breast cancer. 2074:78

98. **TO—VB** (POS bigram, 10,454 instances in corpus)
It is apparent that Charlie did not intend for his haircut *to be* portrayed in an inappropriate manner or for it *to cause* any substantial disruptions. 2074:81

99. **TO—VB—DT** (POS trigram, 3,538 instances in corpus)
It is apparent that Charlie did not intend for his haircut *to be* portrayed in an inappropriate manner or for it *to cause any* substantial disruptions. 2074:81

100. **TO—VB—IN** (POS trigram, 1,373 instances in corpus)
Following the Tinker standard, the Defendant is required *to prove that* the incident in the cafeteria substantially disrupted and interfered with Peanutsberg Junior High School's learning environment in order to justify the suspension and ban of Charlie's haircut. 2074:94

101. **TO—VB—JJ** (POS trigram, 787 instances in corpus)
The District answered, denying the violation of Peterson's free-speech rights and asserting that it had the right *to regulate disruptive* or offensive conduct. 2075:53

102. **TO—VB—NN** (POS trigram, 883 instances in corpus)
The Court ruled in this seminal case that mere fear of a disruption is not enough *to justify abridging* students' speech. 2075:78 (Here, it seems the POS taggers should have tagged "abridging" as VBG instead of as a noun.)
The Court in Fraser granted schools the power *to restrict student* speech in order to promote civility and teach students the boundaries of socially acceptable behavior, in cases of vulgar or plainly offensive speech. 2075:105

103. **TO—VB—NNP** (POS trigram, 341 instances in corpus)
Williams then went on *to tell Witherspoon* that Jacobs had been suspended for a long period of time for the incident, which explained why he was new at Greenburg High School this year. 2078:28

104. **TO—VB—NNS** (POS trigram, 467 instances in corpus)
She is no longer the captain of the cheerleading squad and is not able *to spread rumors* about students with no concern for their wellbeing. 2078:105

105. **TO—VB—Period** (POS trigram, 256 instances in corpus)
The question of whether the privilege was abused must be a question for a jury *to decide.* 2078:121

106. **TO—VB—PRP$** (POS trigram, 433 instances in corpus)
    Williams could have easily told Witherspoon she needs to talk to the school about Jacobs before allowing him *to take her* daughter to the homecoming dance. 2078:133

107. **TO—VB—RB** (POS trigram, 271 instances in corpus)
    Justin told Wager: "I'm going *to get even* with the jerk if it kills me. 2079:40 (It seems that "even" here could have been tagged as an adjective.)
    The contract was created to prevent exactly what Johnson did. 2080:145

108. **TO—VB—VBN** (POS trigram, 916 instances in corpus)
    Morton had *to be transported* by air ambulance to Hennepin County Medical Center and underwent numerous surgeries because of the impact with John Francis. 2081:28

109. **VB—DT—NN** (POS trigram, 4,125 instances in corpus)
    Snow Valley likely uses this policy to recruit and retain employees as well as to *attract the public* to its facility. 2081:42

110. **VB—DT—NNP** (POS trigram, 321 instances in corpus)
    This employee had keys to resident's apartments, and used that access to sexually *assault the Plaintiff*. 2081:96

111. **VB—DT—NNS** (POS trigram, 531 instances in corpus)
    Next, the Plaintiff admits he does not *remember the questions* very well. 2084:118

112. **VB—IN—NNP** (POS trigram, 267 instances in corpus)
    He admits that he does not *remember if Officer* Wright asked him if he had ever been hospitalized for mental health reasons before. 2084:119

113. **VB—TO—VB** (POS trigram, 394 instances in corpus)
    We understand the legislature did not *want to absolve* owners and possessors of property improvements from the duty to use reasonable care to protect health and safety. 2085:61

114. **VB—VBN—TO** (POS trigram, 347 instances in corpus)
    There were no further actions Mr. Smith could *have taken to* prevent this act of God. 2085:71

115. **VBD—DT—NN** (POS trigram, 2,926 instances in corpus)
    In this situation, there *was no reason* for Mr. Smith to believe or foresee that his customers were in risk of any harm when they entered onto his property. 2085:76

116. **VBD—TO—VB** (POS trigram, 1.078 instances in corpus)
    Plaintiff *refused to change* his hair while his sister was undergoing chemotherapy. 2086:5

117. **VBG—DT—NN** (POS trigram, 2,106 instances in corpus)
    When Plaintiff came to school with the words "I Boobies" inscribed in his hair, he was externally *displaying a symbol* for breast cancer awareness. 2086:61

118. **VBN—DT—NN** (POS trigram, 1,398 instances in corpus)
    "Whether a party has primarily *assumed the risk* is usually a question for the jury, unless the evidence is conclusive." 2090:58 (Note that this entire sentence is quoted material.)

119. **VBN—TO—VB** (POS trigram, 943 instances in corpus)
Even if Wager had *tried to involve* himself in Atticus' medical treatment, the doctor's office would not have released Atticus' medical records to Wager for privacy reasons, precisely because Wager was not Atticus' legal guardian. 2091:144

120. **VBP—DT—NN** (POS trigram, 298 instances in corpus)
The Court has determined that school officials *have the authority* to prohibit language that is inappropriate and deters from the educational mission. 2092:75

121. **VBZ—DT** (POS bigram, 3,851 instances in corpus)
It *is a* school's mission to balance these "fundamental values". 2092:102

122. **VBZ—DT—NN** (POS trigram, 2,386 instances in corpus)
It *is a school*'s mission to balance these "fundamental values". 2092:102 (Note that the possessive 's is not included in the featured trigram.)

123. **VBZ—IN—DT** (POS trigram, 1,078 instances in corpus)
The personal grooming and dress policy of Peanutsberg School District specifically *protects against the* valid concerns articulated by the Courts. 2092:121

124. **VBZ—RB—VBN** (POS trigram, 661 instances in corpus)
Atticus' March 21, 2011, threat did not name the decedent and this fact is not disputed. 2093:98

125. **VBZ—TO—VB** (POS trigram, 533 instances in corpus)
The first time somebody *attempts to use* the website in such a way, it will state that all use of storage is subject to the terms and conditions of the website. 2094:21

126. **WDT—DT—NN** (POS trigram, 499 instances in corpus)
In this school, students would more likely know of the breast cancer awareness message *that the phrase* contains. 2098:78

127. **WRB—DT—NN** (POS trigram, 587 instances in corpus)
This becomes critical *when the interpretation* of the Copyright Registry itself is examined. 1001:69

Appendix I

Frequency values for all features in the present study

This appendix consists of tables of the features analyzed in the empirical study in this dissertation. The features are broken into broad categories, one table for each category:

1. Summary table (see Table I.1 beginning at page 288). This table shows only the features of all categories where there was a statistically significant difference in the relative frequency of use by Gender F and Gender M authors.

2. Function words (see Table I.2 beginning at page 291). This category is described in Section 4.3.2, and the list used for this study appears in Appendix C.

3. Parts of speech (see Table I.3 beginning at page 302). This category is described in Section 4.3.2, and the list of POS tags used in this study appears in Appendix B.

4. Part-of-speech bigrams (see Table I.4 beginning at page 304). This category is described in Section 4.3.2, and examples of the bigrams that are discussed in this dissertation drawn from their contexts of use by participants in this study appear in Appendix H.

5. Part-of-speech trigrams (see Table I.5 beginning at page 307). This category is described in Section 4.3.2, and examples of the trigrams that are discussed in this dissertation drawn from their contexts of use by participants in this study appear in Appendix H.

6. Miscellaneous (see Table I.6 beginning at page 322). These features are in two categories: Some were calculated to permit comparison with Argamon, Koppel, Fine, and Shimoni (2003) (see Sections 3.3.2 and 5.2.1 for discussion); others were calculated as a matter of course but not used in the analyses in this dissertation.

I.1 Overview of table contents

Each table presents the following values for each feature in it:

- Gender M's mean frequency of use and standard deviation.
- Gender F's mean frequency of use and standard deviation.
- "Gender prevalence," which indicates which gender used the feature with greater relative frequency on average. *This prevalence should not be interpreted as meaningful unless the Mann–Whitney* p*-value is less than 0.05.*
- Mann-Whitney *p*-value. The difference between Gender F and Gender mean values should be regarded as significant only if the *p*-value is less than 0.05; in each such case, the row is shaded (except for Table I.1, where all the features were significantly different, so no shading is necessary to distinguish them).

Table I.1: **All significantly different features in present study**

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
Function words						
a	0.0261305	0.0047532	0.0246877	0.0050277	M	0.03457
afterward	0.0000000	0.0000000	0.0000118	0.0000532	F	0.03702
all	0.0013621	0.0008766	0.0011060	0.0006484	M	0.03337
by	0.0059910	0.0022564	0.0051846	0.0020064	M	0.00654
can	0.0021423	0.0012278	0.0017606	0.0010884	M	0.02394
either	0.0004332	0.0004409	0.0002741	0.0002791	M	0.01891
for	0.0114260	0.0040204	0.0131429	0.0048829	F	0.01426
furthermore	0.0001862	0.0003369	0.0003258	0.0004601	F	0.00747
herself	0.0000212	0.0000980	0.0000502	0.0001321	F	0.02265
is	0.0127640	0.0034710	0.0117092	0.0032863	M	0.01776
many	0.0003125	0.0005305	0.0001405	0.0002040	M	0.00721
our	0.0001925	0.0003733	0.0001158	0.0003148	M	0.01874
quite	0.0000552	0.0001759	0.0000101	0.0000511	M	0.02394
secondly	0.0000118	0.0000552	0.0000450	0.0001571	F	0.04742
them	0.0008052	0.0006311	0.0006401	0.0006109	M	0.01394
then	0.0006121	0.0005673	0.0004147	0.0004654	M	0.01255
therefore	0.0008398	0.0008367	0.0011499	0.0009779	F	0.00794
today	0.0000079	0.0000430	0.0000380	0.0001347	F	0.04896
under	0.0016216	0.0010742	0.0021334	0.0014687	F	0.01733
whatever	0.0000315	0.0000860	0.0000095	0.0000477	M	0.02884
Parts of speech						
VBZ	0.0251002	0.0049963	0.0228378	0.0045864	M	0.00261
WRB	0.0036779	0.0015932	0.0033209	0.0015325	M	0.03491
POS bigrams						
Comma—NNP	0.0099787	0.0045125	0.0120192	0.0062550	F	0.01615
DT—NNP	0.0268384	0.0136718	0.0229397	0.0143357	M	0.02903
IN—JJ	0.0136721	0.0051510	0.0124783	0.0057618	M	0.04871
NN—Comma	0.0279215	0.0075558	0.0298649	0.0065933	F	0.04541
NN—TO	0.0156157	0.0043774	0.0180115	0.0049224	F	0.00261
NN—VBZ	0.0186459	0.0053610	0.0167931	0.0052461	M	0.00906
NNP—Comma	0.0117854	0.0049543	0.0130815	0.0050589	F	0.03937
PRP—VBZ	0.0059458	0.0029143	0.0051503	0.0028338	M	0.01452
TO—VB	0.0277963	0.0073379	0.0299212	0.0079660	F	0.04754

See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.1: All significantly different features... (continued)

Feature	**Gender M** Mean	**Gender M** Std Dev	**Gender F** Mean	**Gender F** Std Dev	Gender Prevalence	Mann-Whitney p-value
VBD—TO	0.0041960	0.0024843	0.0047787	0.0022601	F	0.02104
VBZ—DT	0.0111129	0.0033163	0.0100883	0.0036115	M	0.02362
VBZ—IN	0.0069723	0.0030081	0.0059624	0.0022746	M	0.01697
VBZ—VBN	0.0082718	0.0033087	0.0072256	0.0028079	M	0.03284
POS trigrams						
Colon—Colon—CD	0.0011753	0.0016144	0.0018115	0.0026544	F	0.04338
Comma—PRP—MD	0.0023817	0.0019281	0.0016966	0.0016598	M	0.01179
DT—JJ—IN	0.0021692	0.0020639	0.0015058	0.0018606	M	0.01014
DT—JJ—NNS	0.0056278	0.0031912	0.0047240	0.0029729	M	0.03986
DT—NN—VBZ	0.0203584	0.0081932	0.0180919	0.0073811	M	0.02809
DT—NNP—VBD	0.0039376	0.0062802	0.0034367	0.0078975	M	0.01379
IN—NNP—NN	0.0012292	0.0017897	0.0017754	0.0021526	F	0.01448
MD—RB—VB	0.0103626	0.0048762	0.0093274	0.0051304	M	0.04639
NN—Comma—NNP	0.0058849	0.0037015	0.0078224	0.0044691	F	0.00100
NN—IN—NNP	0.0176645	0.0083192	0.0210109	0.0093638	F	0.00927
NN—NN—NN	0.0068097	0.0044296	0.0058240	0.0048085	M	0.04203
NN—TO—VB	0.0166442	0.0063993	0.0204396	0.0081413	F	0.00097
NN—VBZ—IN	0.0052461	0.0033614	0.0042501	0.0026341	M	0.03421
NNP—Comma—NNP	0.0051721	0.0041905	0.0068021	0.0052635	F	0.00611
NNP—NNP—Comma	0.0079365	0.0050446	0.0094484	0.0057158	F	0.04541
NNP—NNP—VBZ	0.0060286	0.0041204	0.0048980	0.0036977	M	0.04228
NNP—TO—VB	0.0025544	0.0021852	0.0031398	0.0022159	F	0.03641
NNP—VBD—TO	0.0028764	0.0024726	0.0035553	0.0021121	F	0.00844
NNP—VBZ—IN	0.0029830	0.0023349	0.0022922	0.0021272	M	0.03024
POS—NN—Comma	0.0014146	0.0014316	0.0020867	0.0021449	F	0.03377
POS—NN—TO	0.0018352	0.0018403	0.0025637	0.0023612	F	0.03483
PRP—VBZ—DT	0.0017775	0.0015165	0.0013411	0.0013657	M	0.03600
RB—DT—NN	0.0047540	0.0030972	0.0037187	0.0025519	M	0.02594
TO—DT—NNP	0.0018620	0.0018620	0.0015523	0.0025478	M	0.02148
TO—JJ—NN	0.0011120	0.0018824	0.0016480	0.0023384	F	0.02030
VB—IN—NNP	0.0010833	0.0014275	0.0018256	0.0021117	F	0.01163
VBD—RB—VB	0.0046468	0.0037337	0.0065841	0.0071643	F	0.04428
VBD—TO—VB	0.0051959	0.0040723	0.0066844	0.0037269	F	0.00074
VBZ—DT—NN	0.0137030	0.0047805	0.0126029	0.0053347	M	0.04403

See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.1: All significantly different features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
VBZ—IN—DT	0.0065411	0.0036318	0.0053767	0.0028089	M	0.04126
Miscellaneous						
Pronouns (3rd person plural)	0.0039952	0.0028064	0.0036473	0.0033847	M	0.04151

See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.2: **All function word features in present study**

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
‘d	0.0000000	0.0000000	0.0000050	0.0000507	F	0.36036
‘ll	0.0000728	0.0001294	0.0000495	0.0001242	M	0.11874
‘m	0.0000167	0.0000870	0.0000309	0.0001527	F	0.88188
‘re	0.0000194	0.0000905	0.0000444	0.0001984	F	0.57799
‘s	0.0069029	0.0038214	0.0075043	0.0040462	F	0.22984
‘ve	0.0000068	0.0000455	0.0000023	0.0000232	M	0.46914
a	0.0261305	0.0047532	0.0246877	0.0050277	M	0.03457
about	0.0007747	0.0008017	0.0007477	0.0009519	M	0.28965
above	0.0002382	0.0003583	0.0002179	0.0002810	M	0.79996
according	0.0002391	0.0003376	0.0002769	0.0004322	F	0.73898
accordingly	0.0000927	0.0002022	0.0000989	0.0002262	F	0.79052
actual	0.0002727	0.0004952	0.0002327	0.0004246	M	0.89577
actually	0.0001425	0.0002497	0.0001541	0.0003078	F	0.85356
after	0.0018776	0.0013128	0.0019314	0.0013395	F	0.86249
afterward	0.0000000	0.0000000	0.0000118	0.0000532	F	0.03702
afterwards	0.0000356	0.0001697	0.0000225	0.0000979	M	0.78343
again	0.0001149	0.0002041	0.0001350	0.0002262	F	0.78947
against	0.0006938	0.0006393	0.0006773	0.0006565	M	0.82754
ago	0.0000154	0.0000762	0.0000111	0.0000563	M	0.83399
all	0.0013621	0.0008766	0.0011060	0.0006484	M	0.03337
almost	0.0000509	0.0001218	0.0000571	0.0001464	F	0.92464
along	0.0001025	0.0002038	0.0000534	0.0001337	M	0.08988
already	0.0002211	0.0003567	0.0002649	0.0004257	F	0.53762
also	0.0012783	0.0009258	0.0012568	0.0008276	M	0.95465
although	0.0003200	0.0004799	0.0004095	0.0005160	F	0.10709
always	0.0001202	0.0001929	0.0000889	0.0001974	M	0.09733
am	0.0000056	0.0000532	0.0000103	0.0000620	F	0.40495
among	0.0000996	0.0001757	0.0000713	0.0001600	M	0.16393
an	0.0042780	0.0012872	0.0043498	0.0014790	F	0.87368
and	0.0181814	0.0047875	0.0189160	0.0038259	F	0.10095
another	0.0004343	0.0004828	0.0003814	0.0004268	M	0.58899
any	0.0015903	0.0011156	0.0016735	0.0011811	F	0.67631
anybody	0.0000000	0.0000000	0.0000029	0.0000293	F	0.36036

Findings *significant in this study* ($p < 0.05$) are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.2: All function word features... (continued)

	Gender M		Gender F		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
anyone	0.0000691	0.0002118	0.0001053	0.0003023	F	0.43522
anything	0.0000824	0.0001964	0.0000949	0.0001824	F	0.34393
anywhere	0.0000237	0.0000864	0.0000127	0.0000571	M	0.37544
are	0.0027850	0.0016638	0.0025995	0.0017737	M	0.23489
around	0.0001259	0.0002535	0.0001156	0.0002401	M	0.52898
art	0.0000244	0.0001336	0.0000222	0.0000731	M	0.27623
as	0.0069764	0.0024044	0.0066886	0.0023302	M	0.27474
aside	0.0000151	0.0000746	0.0000136	0.0000616	M	0.92671
at	0.0027546	0.0019404	0.0027999	0.0018971	F	0.50068
away	0.0001435	0.0002207	0.0002244	0.0002943	F	0.05221
back	0.0001085	0.0002140	0.0001193	0.0002416	F	0.82182
be	0.0084048	0.0035254	0.0081319	0.0034533	M	0.56517
bear	0.0000300	0.0000860	0.0000210	0.0000835	M	0.28812
because	0.0026479	0.0017303	0.0025597	0.0014744	M	0.97216
been	0.0011796	0.0008882	0.0011461	0.0008859	M	0.82401
before	0.0015262	0.0015712	0.0015339	0.0012374	F	0.30513
being	0.0006329	0.0007758	0.0004410	0.0004770	M	0.13717
below	0.0003488	0.0004912	0.0003210	0.0003761	M	0.64226
besides	0.0000130	0.0000612	0.0000046	0.0000332	M	0.29222
better	0.0000950	0.0002228	0.0000898	0.0002265	M	0.59797
between	0.0009227	0.0007335	0.0010515	0.0008334	F	0.33014
beyond	0.0001004	0.0002483	0.0000902	0.0002032	M	0.85497
both	0.0013212	0.0012113	0.0011892	0.0009818	M	0.66655
bring	0.0000644	0.0001465	0.0000414	0.0001104	M	0.34605
but	0.0010485	0.0006786	0.0010894	0.0006553	F	0.71162
by	0.0059910	0.0022564	0.0051846	0.0020064	M	0.00654
ca	0.0000132	0.0000727	0.0000050	0.0000507	M	0.25135
came	0.0001185	0.0001884	0.0001512	0.0002035	F	0.22528
can	0.0021423	0.0012278	0.0017606	0.0010884	M	0.02394
certain	0.0002094	0.0002727	0.0002024	0.0002634	M	0.83270
certainly	0.0000902	0.0002336	0.0000349	0.0001026	M	0.12238
come	0.0002207	0.0003971	0.0001532	0.0002716	M	0.41809
comes	0.0000400	0.0001363	0.0000231	0.0000763	M	0.44873
consequently	0.0000528	0.0001753	0.0000681	0.0001685	F	0.25449

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.2: All function word features… (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
could	0.0011095	0.0007661	0.0011827	0.0006685	F	0.30535
definite	0.0001732	0.0002879	0.0001775	0.0002805	F	0.78123
definitely	0.0000000	0.0000000	0.0000054	0.0000385	F	0.19213
despite	0.0001859	0.0002835	0.0001611	0.0002293	M	0.88201
did	0.0017393	0.0011750	0.0022834	0.0023579	F	0.25052
do	0.0005940	0.0004771	0.0005695	0.0004786	M	0.74804
does	0.0013887	0.0008737	0.0013051	0.0008527	M	0.39433
doing	0.0001191	0.0001885	0.0001398	0.0002341	F	0.89542
done	0.0001087	0.0001831	0.0000750	0.0001592	M	0.18917
doubtful	0.0000087	0.0000480	0.0000000	0.0000000	M	0.06065
down	0.0001657	0.0004914	0.0001317	0.0002820	M	0.69707
due	0.0002166	0.0002928	0.0002675	0.0003754	F	0.53998
during	0.0006605	0.0008245	0.0005453	0.0005686	M	0.92125
ePeriodgPeriod	0.0000000	0.0000000	0.0000021	0.0000215	F	0.36036
each	0.0003277	0.0004318	0.0002578	0.0003111	M	0.41660
earlier	0.0001169	0.0002008	0.0000942	0.0001870	M	0.27921
early	0.0000381	0.0001074	0.0000282	0.0000979	M	0.40331
eight	0.0000699	0.0001385	0.0000439	0.0001008	M	0.24875
eighteen	0.0000000	0.0000000	0.0000053	0.0000380	F	0.19213
eighteenth	0.0000060	0.0000564	0.0000029	0.0000297	M	0.91199
eighth	0.0000184	0.0001051	0.0000280	0.0001848	F	0.87003
eighty	0.0000000	0.0000000	0.0000051	0.0000366	F	0.19213
either	0.0004332	0.0004409	0.0002741	0.0002791	M	0.01891
eleven	0.0000111	0.0000635	0.0000123	0.0000553	F	0.63456
eleventh	0.0000173	0.0001633	0.0000241	0.0001885	F	0.65968
else	0.0000526	0.0001261	0.0000306	0.0000944	M	0.20336
enough	0.0001819	0.0002869	0.0002179	0.0003955	F	0.93432
enter	0.0001030	0.0002602	0.0001034	0.0001889	F	0.48704
even	0.0007219	0.0005729	0.0009110	0.0007844	F	0.19400
eventually	0.0000421	0.0001028	0.0000419	0.0000962	M	0.89064
ever	0.0000954	0.0002579	0.0000655	0.0001930	M	0.13688
every	0.0001893	0.0003060	0.0001941	0.0003988	F	0.81043
everyone	0.0000411	0.0001506	0.0000483	0.0001473	F	0.57910
everything	0.0000405	0.0001340	0.0000303	0.0001291	M	0.40792

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.2: All function word features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
example	0.0001692	0.0002360	0.0001736	0.0002549	F	0.89067
except	0.0000222	0.0000935	0.0000359	0.0001067	F	0.25637
exit	0.0000496	0.0002982	0.0000396	0.0002627	M	0.84660
fact	0.0018507	0.0015977	0.0015506	0.0016064	M	0.11260
fair	0.0000984	0.0002228	0.0000952	0.0002166	M	0.81228
far	0.0000705	0.0001515	0.0000612	0.0001398	M	0.68247
few	0.0000825	0.0001581	0.0000945	0.0001962	F	1.00000
fewer	0.0000028	0.0000265	0.0000075	0.0000447	F	0.39907
fifteen	0.0000544	0.0001125	0.0000736	0.0001535	F	0.49672
fifteenth	0.0000029	0.0000278	0.0000000	0.0000000	M	0.28435
fifth	0.0000457	0.0001443	0.0000604	0.0001681	F	0.41779
fifty	0.0000174	0.0000918	0.0000071	0.0000414	M	0.54414
finally	0.0001872	0.0002627	0.0001631	0.0002302	M	0.66647
first	0.0010909	0.0009512	0.0011045	0.0009706	F	0.99484
firstly	0.0000033	0.0000316	0.0000055	0.0000402	F	0.65968
five	0.0001396	0.0002465	0.0001159	0.0002272	M	0.34396
for	0.0114260	0.0040204	0.0131429	0.0048829	F	0.01426
forever	0.0000135	0.0000751	0.0000000	0.0000000	M	0.06065
forgo	0.0000065	0.0000438	0.0000000	0.0000000	M	0.12717
forth	0.0001065	0.0002433	0.0001278	0.0002051	F	0.10302
forty	0.0000137	0.0000779	0.0000048	0.0000343	M	0.52601
four	0.0001401	0.0002853	0.0001245	0.0002528	M	0.77754
fourteen	0.0000032	0.0000299	0.0000029	0.0000295	M	0.92368
fourteenth	0.0000160	0.0000768	0.0000027	0.0000278	M	0.12340
fourth	0.0000300	0.0001103	0.0000360	0.0001808	F	0.43263
from	0.0027387	0.0015103	0.0027261	0.0015296	M	0.82809
furthermore	0.0001862	0.0003369	0.0003258	0.0004601	F	0.00747
generally	0.0001698	0.0002238	0.0001366	0.0002474	M	0.08579
get	0.0001481	0.0002667	0.0002862	0.0005273	F	0.21015
getting	0.0000461	0.0001973	0.0000460	0.0001372	M	0.41071
give	0.0001693	0.0002615	0.0002537	0.0003273	F	0.07739
go	0.0001189	0.0002269	0.0000991	0.0002123	M	0.43256
good	0.0001608	0.0003178	0.0001168	0.0002470	M	0.35165
got	0.0000269	0.0001199	0.0000795	0.0002463	F	0.11726

Findings *significant in this study* (*p* < 0.05) are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.2: All function word features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
had	0.0019677	0.0014846	0.0022410	0.0015776	F	0.17186
has	0.0025669	0.0015553	0.0023240	0.0013050	M	0.40957
have	0.0024895	0.0013071	0.0025682	0.0012679	F	0.44956
having	0.0001843	0.0002673	0.0001468	0.0002784	M	0.31487
he	0.0023880	0.0030703	0.0025901	0.0031893	F	0.37850
hence	0.0000091	0.0000497	0.0000238	0.0001012	F	0.29841
her	0.0018658	0.0026531	0.0019323	0.0021078	F	0.25140
here	0.0004345	0.0005620	0.0004135	0.0005498	M	0.72072
hers	0.0000031	0.0000296	0.0000024	0.0000249	M	0.91199
herself	0.0000212	0.0000980	0.0000502	0.0001321	F	0.02265
him	0.0005970	0.0009237	0.0005459	0.0008786	M	0.55476
himself	0.0001311	0.0003280	0.0001473	0.0003496	F	0.59693
his	0.0026278	0.0032498	0.0026256	0.0033424	M	0.62694
how	0.0002541	0.0003408	0.0002115	0.0003194	M	0.29736
however	0.0008637	0.0007008	0.0010039	0.0006707	F	0.10737
hundred	0.0000030	0.0000286	0.0000077	0.0000450	F	0.39324
i	0.0004260	0.0006129	0.0005009	0.0007427	F	0.53312
if	0.0029075	0.0013747	0.0026517	0.0011729	M	0.25861
in	0.0168977	0.0034287	0.0165344	0.0032332	M	0.44647
indeed	0.0000686	0.0001725	0.0000615	0.0001613	M	0.98899
instance	0.0000879	0.0001964	0.0000652	0.0001460	M	0.54103
instead	0.0002308	0.0003123	0.0002165	0.0002961	M	0.91437
into	0.0008409	0.0007515	0.0009822	0.0008178	F	0.26424
is	0.0127640	0.0034710	0.0117092	0.0032863	M	0.01776
it	0.0056206	0.0023927	0.0049424	0.0020420	M	0.06055
its	0.0016085	0.0010751	0.0016907	0.0011392	F	0.60782
itself	0.0002141	0.0002999	0.0001775	0.0002800	M	0.28860
last	0.0000831	0.0002235	0.0001004	0.0003135	F	0.97117
lastly	0.0000165	0.0000824	0.0000304	0.0000923	F	0.12706
later	0.0005254	0.0005940	0.0004621	0.0004741	M	0.75278
less	0.0001218	0.0002486	0.0001369	0.0001916	F	0.11912
let	0.0000258	0.0000940	0.0000381	0.0001258	F	0.39803
like	0.0002644	0.0003651	0.0003625	0.0005531	F	0.22558
likely	0.0002112	0.0003170	0.0002236	0.0003443	F	0.97956

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.2: All function word features… (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
many	0.0003125	0.0005305	0.0001405	0.0002040	M	0.00721
matter	0.0007162	0.0007394	0.0006784	0.0006739	M	0.88943
may	0.0011788	0.0008704	0.0011072	0.0007824	M	0.76917
maybe	0.0000172	0.0000831	0.0000212	0.0000948	F	0.71376
me	0.0000451	0.0001973	0.0000866	0.0002911	F	0.19012
might	0.0002086	0.0003643	0.0001550	0.0002370	M	0.62044
million	0.0000339	0.0002459	0.0000085	0.0000690	M	0.86277
mine	0.0000048	0.0000456	0.0000000	0.0000000	M	0.28435
more	0.0008805	0.0006723	0.0008430	0.0005991	M	0.90429
moreover	0.0001304	0.0003413	0.0000881	0.0002337	M	0.56298
most	0.0004676	0.0005128	0.0004146	0.0003908	M	0.88486
much	0.0001460	0.0002178	0.0001787	0.0002965	F	0.69004
must	0.0021069	0.0012294	0.0023056	0.0013322	F	0.40591
my	0.0000119	0.0000687	0.0000194	0.0001002	F	0.61338
myself	0.0000031	0.0000296	0.0000031	0.0000312	M	0.92368
n’	0.0001929	0.0005677	0.0001367	0.0003219	M	0.85025
near	0.0000717	0.0003348	0.0000515	0.0001818	M	0.84339
nearby	0.0000223	0.0001057	0.0000168	0.0001220	M	0.54414
nearly	0.0000250	0.0000856	0.0000549	0.0001673	F	0.28650
neither	0.0001736	0.0002736	0.0001440	0.0002345	M	0.51951
never	0.0003542	0.0003670	0.0003977	0.0005832	F	0.45967
nevertheless	0.0000244	0.0000886	0.0000311	0.0001300	F	0.82932
next	0.0000807	0.0001378	0.0000966	0.0001840	F	0.92796
nine	0.0000654	0.0001911	0.0001019	0.0002339	F	0.14055
nineteen	0.0000032	0.0000300	0.0000028	0.0000285	M	0.92368
ninth	0.0001248	0.0003284	0.0001015	0.0003111	M	0.50412
no	0.0021428	0.0012858	0.0022340	0.0014142	F	0.70010
nobody	0.0000083	0.0000787	0.0000027	0.0000278	M	0.91199
none	0.0000539	0.0001307	0.0000308	0.0000965	M	0.19274
nor	0.0002428	0.0003876	0.0002530	0.0003631	F	0.65289
not	0.0089641	0.0028401	0.0094287	0.0040508	F	0.89716
nothing	0.0001307	0.0002175	0.0001217	0.0002113	M	0.74111
now	0.0001312	0.0001920	0.0001552	0.0002133	F	0.47540
nowhere	0.0000315	0.0001204	0.0000235	0.0000860	M	0.94008

Findings *significant in this study* ($p < 0.05$) are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.2: All function word features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
o	0.0000139	0.0000648	0.0000140	0.0000631	F	0.93796
occasionally	0.0000000	0.0000000	0.0000027	0.0000275	F	0.36036
of	0.0323818	0.0056809	0.0317297	0.0055996	M	0.37385
off	0.0001649	0.0004911	0.0001261	0.0003781	M	0.64696
often	0.0000633	0.0001401	0.0000778	0.0001594	F	0.63071
on	0.0059237	0.0019919	0.0058563	0.0021897	M	0.60424
once	0.0001455	0.0002001	0.0001026	0.0001936	M	0.06247
one	0.0015160	0.0009761	0.0013604	0.0008910	M	0.28335
only	0.0014363	0.0007721	0.0013113	0.0007600	M	0.35472
or	0.0043621	0.0021486	0.0042775	0.0024302	M	0.67348
order	0.0008267	0.0006379	0.0010641	0.0009055	F	0.14650
other	0.0012016	0.0007429	0.0011459	0.0007750	M	0.37384
others	0.0001698	0.0003259	0.0001979	0.0003905	F	0.92450
ought	0.0000432	0.0001452	0.0000129	0.0000587	M	0.09759
our	0.0001925	0.0003733	0.0001158	0.0003148	M	0.01874
ours	0.0000031	0.0000288	0.0000000	0.0000000	M	0.28435
ourselves	0.0000000	0.0000000	0.0000021	0.0000211	F	0.36036
out	0.0004935	0.0006016	0.0006216	0.0006777	F	0.12913
over	0.0004540	0.0003992	0.0004362	0.0004966	M	0.33172
perhaps	0.0000484	0.0001294	0.0000309	0.0001216	M	0.13631
possible	0.0001286	0.0002033	0.0001615	0.0002462	F	0.55163
possibly	0.0000335	0.0000910	0.0000260	0.0000954	M	0.31844
presumably	0.0000230	0.0000832	0.0000131	0.0000629	M	0.34941
previous	0.0001639	0.0003102	0.0001864	0.0003098	F	0.35826
previously	0.0000960	0.0001679	0.0001092	0.0001954	F	0.91651
prior	0.0014219	0.0018053	0.0015394	0.0017660	F	0.66653
probably	0.0000450	0.0001543	0.0000362	0.0001069	M	0.87768
quite	0.0000552	0.0001759	0.0000101	0.0000511	M	0.02394
rare	0.0000030	0.0000284	0.0000122	0.0000548	F	0.14501
rarely	0.0000142	0.0000660	0.0000139	0.0000626	M	0.95486
rather	0.0003289	0.0004231	0.0002567	0.0003356	M	0.35544
result	0.0005553	0.0006177	0.0004516	0.0005322	M	0.15460
resulting	0.0000938	0.0002018	0.0000712	0.0001725	M	0.32732
round	0.0000000	0.0000000	0.0000055	0.0000397	F	0.19213

Findings *significant in this study* ($p < 0.05$) are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.2: All function word features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
said	0.0002854	0.0004081	0.0002623	0.0004617	M	0.41940
same	0.0004177	0.0004245	0.0004446	0.0004649	F	0.84120
say	0.0000950	0.0002254	0.0001201	0.0002108	F	0.21570
second	0.0009861	0.0010266	0.0008960	0.0009033	M	0.92151
secondly	0.0000118	0.0000552	0.0000450	0.0001571	F	0.04742
seldom	0.0000000	0.0000000	0.0000055	0.0000394	F	0.19213
seven	0.0000648	0.0003892	0.0000106	0.0000533	M	0.23519
seventeen	0.0000000	0.0000000	0.0000030	0.0000302	F	0.36036
seventh	0.0003666	0.0005733	0.0003507	0.0005904	M	0.91394
seventy	0.0000024	0.0000226	0.0000000	0.0000000	M	0.28435
shall	0.0006400	0.0007464	0.0005882	0.0007727	M	0.41175
she	0.0015423	0.0020874	0.0017867	0.0024098	F	0.44793
should	0.0028236	0.0017684	0.0026250	0.0015299	M	0.61601
similarly	0.0001107	0.0002373	0.0001365	0.0002483	F	0.32154
since	0.0004694	0.0005284	0.0004690	0.0005783	M	0.73351
six	0.0000258	0.0000774	0.0000392	0.0001474	F	0.79186
sixteen	0.0000128	0.0000732	0.0000000	0.0000000	M	0.06065
sixty	0.0000028	0.0000266	0.0000000	0.0000000	M	0.28435
so	0.0008542	0.0007519	0.0007631	0.0005363	M	0.78894
some	0.0005042	0.0005042	0.0004022	0.0003937	M	0.19290
somebody	0.0000066	0.0000440	0.0000000	0.0000000	M	0.12717
someone	0.0000908	0.0001703	0.0000622	0.0001562	M	0.09960
something	0.0001656	0.0002770	0.0001234	0.0002308	M	0.25001
sometimes	0.0000142	0.0000673	0.0000328	0.0001137	F	0.18136
somewhere	0.0000000	0.0000000	0.0000027	0.0000278	F	0.36036
soon	0.0000515	0.0001309	0.0000250	0.0000779	M	0.18414
still	0.0002825	0.0003204	0.0003020	0.0003583	F	0.80199
subsequently	0.0001105	0.0002177	0.0000862	0.0001560	M	0.91787
such	0.0014267	0.0009193	0.0012643	0.0011216	M	0.05486
sure	0.0000236	0.0000869	0.0000397	0.0001038	F	0.20590
tell	0.0000273	0.0001009	0.0000464	0.0001482	F	0.29902
ten	0.0000314	0.0001020	0.0000565	0.0001613	F	0.28066
than	0.0007145	0.0005033	0.0006886	0.0004654	M	0.59331
that	0.0175249	0.0046036	0.0170830	0.0047702	M	0.78606

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.2: All function word features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
the	0.0754758	0.0126603	0.0739097	0.0163666	M	0.60424
their	0.0015565	0.0012339	0.0015314	0.0018454	M	0.15356
theirs	0.0000029	0.0000277	0.0000026	0.0000268	M	0.92368
them	0.0008052	0.0006311	0.0006401	0.0006109	M	0.01394
themselves	0.0000811	0.0001914	0.0000669	0.0001496	M	0.78863
then	0.0006121	0.0005673	0.0004147	0.0004654	M	0.01255
there	0.0026811	0.0017689	0.0025173	0.0015166	M	0.61147
therefore	0.0008398	0.0008367	0.0011499	0.0009779	F	0.00794
these	0.0010886	0.0010329	0.0009002	0.0007877	M	0.25172
they	0.0015494	0.0014166	0.0014064	0.0015178	M	0.15320
third	0.0001534	0.0002325	0.0001920	0.0003384	F	0.85526
thirdly	0.0000064	0.0000425	0.0000053	0.0000390	M	0.87927
thirteen	0.0000028	0.0000260	0.0000050	0.0000507	F	0.92368
thirty	0.0000126	0.0000590	0.0000275	0.0000949	F	0.24717
this	0.0060020	0.0024778	0.0055976	0.0026260	M	0.17146
those	0.0004440	0.0004236	0.0004013	0.0004283	M	0.38565
though	0.0002451	0.0003611	0.0002303	0.0003222	M	0.61537
thousand	0.0000035	0.0000334	0.0000000	0.0000000	M	0.28435
three	0.0002365	0.0003700	0.0002390	0.0003048	F	0.72471
through	0.0004613	0.0006595	0.0003950	0.0004438	M	0.59441
thus	0.0006792	0.0008890	0.0006099	0.0008103	M	0.73406
to	0.0267082	0.0046815	0.0278094	0.0049684	F	0.08350
today	0.0000079	0.0000430	0.0000380	0.0001347	F	0.04896
tomorrow	0.0000000	0.0000000	0.0000024	0.0000249	F	0.36036
too	0.0001263	0.0001988	0.0001426	0.0002491	F	0.98293
towards	0.0000578	0.0001784	0.0000674	0.0001772	F	0.43740
twelve	0.0000123	0.0000580	0.0000174	0.0000657	F	0.51996
twentieth	0.0000125	0.0001178	0.0000000	0.0000000	M	0.28435
twenty	0.0000602	0.0002622	0.0000172	0.0000729	M	0.10644
twice	0.0000094	0.0000507	0.0000048	0.0000347	M	0.52601
two	0.0006880	0.0005594	0.0006863	0.0006111	M	0.82456
under	0.0016216	0.0010742	0.0021334	0.0014687	F	0.01733
undergo	0.0000029	0.0000271	0.0000000	0.0000000	M	0.28435
undoubtedly	0.0000146	0.0000603	0.0000121	0.0000635	M	0.58101

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.2: All function word features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
unless	0.0002584	0.0002972	0.0003096	0.0003068	F	0.21199
unlikely	0.0000415	0.0001201	0.0000522	0.0001232	F	0.37575
until	0.0003046	0.0004249	0.0003400	0.0003883	F	0.35877
unto	0.0000033	0.0000316	0.0000000	0.0000000	M	0.28435
unusual	0.0000153	0.0000638	0.0000277	0.0000944	F	0.39647
unusually	0.0000000	0.0000000	0.0000024	0.0000248	F	0.36036
up	0.0003723	0.0005013	0.0002596	0.0003453	M	0.13168
upon	0.0006594	0.0006529	0.0005971	0.0005549	M	0.44859
us	0.0000504	0.0001600	0.0000271	0.0000857	M	0.30559
very	0.0002242	0.0002752	0.0002403	0.0003988	F	0.71216
was	0.0082049	0.0033351	0.0081241	0.0041023	M	0.44109
way	0.0003530	0.0003708	0.0003186	0.0003564	M	0.38490
we	0.0003408	0.0004242	0.0003170	0.0004913	M	0.36888
well	0.0003979	0.0004302	0.0003589	0.0004325	M	0.35076
were	0.0019687	0.0014171	0.0019008	0.0015476	M	0.54695
what	0.0008163	0.0007530	0.0006440	0.0006270	M	0.12089
whatever	0.0000315	0.0000860	0.0000095	0.0000477	M	0.02884
when	0.0024160	0.0013157	0.0023414	0.0011962	M	0.67726
where	0.0007992	0.0007092	0.0006434	0.0005362	M	0.22597
whereas	0.0000530	0.0001238	0.0000467	0.0001227	M	0.56033
wherefore	0.0000036	0.0000338	0.0000000	0.0000000	M	0.28435
whether	0.0011638	0.0011691	0.0009670	0.0010506	M	0.21242
which	0.0023446	0.0014383	0.0022304	0.0013508	M	0.66783
while	0.0008274	0.0010394	0.0007921	0.0007063	M	0.35228
who	0.0007969	0.0005170	0.0009689	0.0008038	F	0.39331
whoever	0.0000022	0.0000211	0.0000026	0.0000264	F	0.92368
whom	0.0001313	0.0001942	0.0001216	0.0002090	M	0.39506
whose	0.0000584	0.0001303	0.0000387	0.0000974	M	0.33912
why	0.0001878	0.0002796	0.0001057	0.0001751	M	0.05939
will	0.0009740	0.0007212	0.0009759	0.0007295	F	0.94948
with	0.0040157	0.0014270	0.0040749	0.0018600	F	0.52482
within	0.0006229	0.0009232	0.0006842	0.0008394	F	0.59597
without	0.0005597	0.0004843	0.0005808	0.0005353	F	0.96282
would	0.0034694	0.0019228	0.0031957	0.0016159	M	0.40518

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.2: All function word features… (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
yes	0.0001186	0.0003113	0.0001018	0.0003029	M	0.50544
yesterday	0.0000000	0.0000000	0.0000042	0.0000429	F	0.36036
yet	0.0001866	0.0003043	0.0001658	0.0003063	M	0.52945
you	0.0002058	0.0003424	0.0002318	0.0005615	F	0.52580
your	0.0001001	0.0001742	0.0001194	0.0003620	F	0.55245

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.3: **All part-of-speech features in present study**

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
Comma	0.0374189	0.0083051	0.0393302	0.0068906	F	0.12688
Colon	0.0060488	0.0035828	0.0057972	0.0037581	M	0.53839
Period	0.0391738	0.0059562	0.0383638	0.0057792	M	0.25698
$	0.0003384	0.0003837	0.0003380	0.0003812	M	0.85547
OQuote	0.0124679	0.0060089	0.0135473	0.0068950	F	0.39361
CQuote	0.0052939	0.0037757	0.0057934	0.0042134	F	0.47315
CC	0.0245062	0.0057249	0.0251591	0.0052332	F	0.34473
CD	0.0164519	0.0109170	0.0175303	0.0105717	F	0.34605
DT	0.1226506	0.0133626	0.1184478	0.0168323	M	0.09016
EX	0.0024531	0.0016570	0.0023355	0.0014141	M	0.79801
FW	0.0000030	0.0000286	0.0000029	0.0000297	M	0.92368
IN	0.1259663	0.0086297	0.1253076	0.0108034	M	0.85945
JJ	0.0506543	0.0083568	0.0494074	0.0098780	M	0.19796
JJR	0.0011691	0.0007435	0.0010669	0.0007146	M	0.15582
JJS	0.0007427	0.0006216	0.0006359	0.0005725	M	0.16237
LS	0.0001266	0.0002661	0.0001160	0.0003438	M	0.25715
MD	0.0144645	0.0040998	0.0136705	0.0038497	M	0.15247
NN	0.1734450	0.0135938	0.1742111	0.0145492	F	0.77911
NNP	0.0716031	0.0188542	0.0747782	0.0213467	F	0.36764
NNPS	0.0007260	0.0007223	0.0007868	0.0007543	F	0.65900
NNS	0.0431726	0.0086439	0.0419030	0.0091115	M	0.43726
PDT	0.0000503	0.0001345	0.0000326	0.0000937	M	0.40293
POS	0.0085737	0.0037933	0.0092974	0.0043806	F	0.31647
PRP	0.0143970	0.0063856	0.0139744	0.0062980	M	0.50480
PRP$	0.0076150	0.0040962	0.0075755	0.0039656	M	0.99381
RB	0.0348833	0.0068577	0.0349685	0.0072421	F	0.79801
RBR	0.0005487	0.0004770	0.0005731	0.0005191	F	0.93900
RBS	0.0001023	0.0002174	0.0001222	0.0002261	F	0.55691
RP	0.0008915	0.0008249	0.0008565	0.0007455	M	0.88993
TO	0.0266912	0.0046856	0.0278046	0.0049673	F	0.07943
UH	0.0000054	0.0000357	0.0000129	0.0000584	F	0.34223
VB	0.0327341	0.0063571	0.0335852	0.0053776	F	0.18780
VBD	0.0317105	0.0092527	0.0338566	0.0093981	F	0.11158

Findings *significant in this study* ($p < 0.05$) are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.3: All part-of-speech features… (continued)

	Gender M		Gender F		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
VBG	0.0138753	0.0039590	0.0127895	0.0035788	M	0.07426
VBN	0.0349974	0.0087331	0.0349552	0.0094846	M	0.92277
VBP	0.0074573	0.0026591	0.0073695	0.0030269	M	0.59436
VBZ	0.0251002	0.0049963	0.0228378	0.0045864	M	0.00261
WDT	0.0057962	0.0017239	0.0056127	0.0019400	M	0.54781
WP	0.0017191	0.0008950	0.0017071	0.0011345	M	0.57042
WP$	0.0000584	0.0001303	0.0000387	0.0000974	M	0.33912
WRB	0.0036779	0.0015932	0.0033209	0.0015325	M	0.03491

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.4: **All POS bigram features in present study**

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
Comma—CC	0.0095864	0.0044124	0.0096555	0.0044043	F	0.85133
Comma—DT	0.0138505	0.0047493	0.0146457	0.0044403	F	0.22984
Comma—IN	0.0082048	0.0036814	0.0077713	0.0035344	M	0.34539
Comma—NNP	0.0099787	0.0045125	0.0120192	0.0062550	F	0.01615
OQuote—DT	0.0049245	0.0027295	0.0048635	0.0031440	M	0.64633
CC—DT	0.0050539	0.0023641	0.0046650	0.0020380	M	0.32150
CC—JJ	0.0045716	0.0029073	0.0048268	0.0028450	F	0.43422
CC—NN	0.0076030	0.0040605	0.0073358	0.0033601	M	0.91251
CC—NNP	0.0051985	0.0033076	0.0060886	0.0039528	F	0.15471
CD—CD	0.0080127	0.0112859	0.0088611	0.0096471	F	0.20103
CD—Comma	0.0062620	0.0040424	0.0062799	0.0039034	F	0.93304
DT—JJ	0.0338492	0.0070458	0.0324021	0.0070731	M	0.35338
DT—NN	0.1413730	0.0224295	0.1402914	0.0246747	M	0.76921
DT—NNP	0.0268384	0.0136718	0.0229397	0.0143357	M	0.02903
DT—NNS	0.0208226	0.0068596	0.0201073	0.0079408	M	0.42441
DT—VBN	0.0076240	0.0066314	0.0076467	0.0058645	F	0.83312
IN—CD	0.0054957	0.0039002	0.0052014	0.0035888	M	0.59167
IN—DT	0.1072204	0.0153713	0.1047795	0.0192311	M	0.46521
IN—JJ	0.0136721	0.0051510	0.0124783	0.0057618	M	0.04871
IN—NN	0.0437590	0.0109486	0.0460049	0.0096073	F	0.06325
IN—NNP	0.0282134	0.0092803	0.0306555	0.0112398	F	0.23185
IN—NNS	0.0077191	0.0032571	0.0076558	0.0032968	M	0.81803
IN—PRP	0.0086641	0.0043787	0.0086501	0.0047545	M	0.68770
IN—PRP$	0.0076439	0.0043050	0.0074591	0.0047596	M	0.42441
IN—VBG	0.0056257	0.0029862	0.0050074	0.0021950	M	0.35742
JJ—Comma	0.0043713	0.0020841	0.0044016	0.0020751	F	0.98041
JJ—IN	0.0103922	0.0045131	0.0097985	0.0039688	M	0.32466
JJ—NN	0.0452028	0.0089271	0.0449993	0.0093458	M	0.85742
JJ—NNS	0.0138040	0.0046992	0.0129468	0.0052667	M	0.10230
JJ—Period	0.0061564	0.0035487	0.0062903	0.0033024	F	0.73780
JJ—TO	0.0057085	0.0030194	0.0057403	0.0032411	F	0.96907
MD—RB	0.0056628	0.0026111	0.0052454	0.0027361	M	0.14916
MD—VB	0.0219571	0.0066287	0.0209949	0.0065179	M	0.32339

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.4: All POS bigram features… (continued)

Feature	Gender M Mean	Gender M Std Dev	Gender F Mean	Gender F Std Dev	Gender Prevalence	Mann-Whitney p-value
NN—CC	0.0149967	0.0053053	0.0156111	0.0049474	F	0.38507
NN—Colon	0.0050021	0.0032790	0.0047612	0.0038283	M	0.44185
NN—Comma	0.0279215	0.0075558	0.0298649	0.0065933	F	0.04541
NN—DT	0.0075644	0.0028260	0.0069734	0.0026881	M	0.13957
NN—IN	0.0961915	0.0122826	0.0974480	0.0131385	F	0.44416
NN—MD	0.0110725	0.0044118	0.0107788	0.0043484	M	0.55993
NN—NN	0.0395631	0.0098468	0.0385762	0.0108152	M	0.43270
NN—NNS	0.0119929	0.0050847	0.0118928	0.0046007	M	0.97216
NN—Period	0.0372823	0.0073612	0.0369062	0.0081473	M	0.43422
NN—RB	0.0052433	0.0023979	0.0051863	0.0022180	M	0.76921
NN—TO	0.0156157	0.0043774	0.0180115	0.0049224	F	0.00261
NN—VBD	0.0168520	0.0064849	0.0160665	0.0051293	M	0.66689
NN—VBG	0.0043723	0.0021887	0.0039453	0.0020232	M	0.16586
NN—VBN	0.0071273	0.0064628	0.0077645	0.0064144	F	0.37524
NN—VBZ	0.0186459	0.0053610	0.0167931	0.0052461	M	0.00906
NN—WDT	0.0042835	0.0018875	0.0039389	0.0019161	M	0.23286
NNP—CC	0.0057776	0.0033633	0.0064600	0.0035526	F	0.11782
NNP—CD	0.0060930	0.0048324	0.0065671	0.0043294	F	0.24210
NNP—Comma	0.0117854	0.0049543	0.0130815	0.0050589	F	0.03937
NNP—IN	0.0110816	0.0047993	0.0105948	0.0054111	M	0.21610
NNP—MD	0.0045879	0.0033583	0.0043092	0.0026072	M	0.84323
NNP—NN	0.0057125	0.0037551	0.0055996	0.0032839	M	0.78804
NNP—NNP	0.0379815	0.0193587	0.0380722	0.0192426	F	0.96598
NNP—Period	0.0080434	0.0037368	0.0085506	0.0040069	F	0.35809
NNP—POS	0.0107914	0.0061661	0.0118878	0.0069830	F	0.41031
NNP—VBD	0.0175527	0.0065165	0.0205722	0.0106707	F	0.10095
NNP—VBZ	0.0085865	0.0038045	0.0080111	0.0035897	M	0.27989
NNS—CC	0.0054501	0.0045394	0.0049747	0.0029525	M	0.96392
NNS—Comma	0.0077727	0.0033959	0.0079162	0.0032400	F	0.57042
NNS—IN	0.0221289	0.0057723	0.0212327	0.0060823	M	0.36832
NNS—Period	0.0089665	0.0037952	0.0086583	0.0036354	M	0.54438
NNS—TO	0.0044158	0.0020800	0.0041317	0.0019239	M	0.31961
NNS—VBD	0.0057157	0.0031699	0.0056586	0.0028965	M	0.88694
NNS—VBP	0.0059209	0.0028237	0.0060150	0.0032024	F	0.94744

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.4: All POS bigram features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
POS—NN	0.0106145	0.0049286	0.0116015	0.0063712	F	0.38014
PRP—VBD	0.0082331	0.0064955	0.0090633	0.0064150	F	0.17514
PRP—VBZ	0.0059458	0.0029143	0.0051503	0.0028338	M	0.01452
PRP$—NN	0.0095080	0.0047995	0.0096077	0.0050761	F	0.97938
RB—Comma	0.0070277	0.0034459	0.0075835	0.0035490	F	0.33817
RB—IN	0.0068149	0.0026881	0.0070024	0.0023248	F	0.34605
RB—JJ	0.0063981	0.0034159	0.0060029	0.0033701	M	0.43194
RB—VB	0.0114816	0.0039335	0.0120603	0.0054176	F	0.93099
RB—VBD	0.0041487	0.0022086	0.0042977	0.0021356	F	0.57835
RB—VBN	0.0087552	0.0032545	0.0088493	0.0028769	F	0.82809
TO—DT	0.0109504	0.0035728	0.0109917	0.0035103	F	0.75149
TO—VB	0.0277963	0.0073379	0.0299212	0.0079660	F	0.04754
VB—DT	0.0181144	0.0052797	0.0188649	0.0049842	F	0.30904
VB—IN	0.0082243	0.0030696	0.0088500	0.0030105	F	0.22093
VB—JJ	0.0046341	0.0023853	0.0045515	0.0021625	M	0.85742
VB—VBN	0.0131231	0.0053395	0.0130637	0.0058787	M	0.78506
VBD—DT	0.0115954	0.0042844	0.0120268	0.0042194	F	0.45813
VBD—IN	0.0126904	0.0044263	0.0133678	0.0047636	F	0.24418
VBD—RB	0.0087470	0.0042539	0.0100027	0.0066920	F	0.45501
VBD—TO	0.0041960	0.0024843	0.0047787	0.0022601	F	0.02104
VBD—VBN	0.0085669	0.0037604	0.0083834	0.0037652	M	0.70202
VBG—DT	0.0077594	0.0029368	0.0070932	0.0030794	M	0.05363
VBG—IN	0.0046590	0.0024878	0.0042467	0.0027541	M	0.07943
VBN—DT	0.0061575	0.0023995	0.0055566	0.0024355	M	0.06216
VBN—IN	0.0258707	0.0088138	0.0269954	0.0099392	F	0.45735
VBN—NN	0.0093270	0.0069660	0.0094274	0.0064515	F	0.70586
VBN—Period	0.0040678	0.0019748	0.0040873	0.0022253	F	0.97938
VBN—TO	0.0056229	0.0029682	0.0052706	0.0026400	M	0.41769
VBZ—DT	0.0111129	0.0033163	0.0100883	0.0036115	M	0.02362
VBZ—IN	0.0069723	0.0030081	0.0059624	0.0022746	M	0.01697
VBZ—JJ	0.0055292	0.0026048	0.0056282	0.0023654	F	0.54181
VBZ—RB	0.0089022	0.0035393	0.0082739	0.0032366	M	0.19438
VBZ—VBN	0.0082718	0.0033087	0.0072256	0.0028079	M	0.03284

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: **All POS trigram features in present study**

Feature	Gender M Mean	Gender M Std Dev	Gender F Mean	Gender F Std Dev	Gender Prevalence	Mann-Whitney p-value
Colon—CD—CD	0.0023589	0.0036754	0.0033972	0.0051186	F	0.05639
Colon—Colon—CD	0.0011753	0.0016144	0.0018115	0.0026544	F	0.04338
Colon—DT—NN	0.0014173	0.0017370	0.0012505	0.0013915	M	0.80794
Comma—CC—DT	0.0030682	0.0022740	0.0029670	0.0025152	M	0.72670
Comma—CC—IN	0.0021563	0.0020840	0.0023446	0.0022074	F	0.52030
Comma—CC—NN	0.0015408	0.0016716	0.0013423	0.0015301	M	0.40801
Comma—CC—NNS	0.0014185	0.0016218	0.0014714	0.0012990	F	0.31329
Comma—CC—RB	0.0023598	0.0022062	0.0021415	0.0019784	M	0.57412
Comma—CC—VBD	0.0016036	0.0016584	0.0016186	0.0015019	F	0.62639
Comma—CD—Comma	0.0037414	0.0029164	0.0033892	0.0028952	M	0.36622
Comma—DT—JJ	0.0028705	0.0018683	0.0030927	0.0021945	F	0.66959
Comma—DT—NN	0.0140609	0.0055345	0.0155298	0.0054704	F	0.07681
Comma—DT—NNP	0.0063634	0.0050184	0.0059779	0.0056382	M	0.36268
Comma—DT—NNS	0.0021760	0.0019745	0.0024977	0.0019129	F	0.22628
Comma—IN—DT	0.0062035	0.0035789	0.0058653	0.0041588	M	0.21996
Comma—IN—NN	0.0025276	0.0019952	0.0023097	0.0017288	M	0.63587
Comma—IN—NNP	0.0014334	0.0014013	0.0017017	0.0016176	F	0.26086
Comma—JJ—NN	0.0024725	0.0022119	0.0023966	0.0018773	M	0.87535
Comma—NN—Comma	0.0010361	0.0013774	0.0017833	0.0027491	F	0.12804
Comma—NNP—CC	0.0014487	0.0018703	0.0015682	0.0017271	F	0.33959
Comma—NNP—NNP	0.0062349	0.0044004	0.0074010	0.0067699	F	0.41177
Comma—NNP—POS	0.0012769	0.0018366	0.0016158	0.0020599	F	0.09187
Comma—NNP—VBD	0.0047931	0.0037270	0.0062325	0.0055676	F	0.14225
Comma—NNP—VBZ	0.0017765	0.0024133	0.0017136	0.0018407	M	0.80148
Comma—NNS—Comma	0.0014159	0.0019842	0.0014913	0.0018342	F	0.50009
Comma—PRP—MD	0.0023817	0.0019281	0.0016966	0.0016598	M	0.01179
Comma—PRP—VBD	0.0017619	0.0021362	0.0019034	0.0021122	F	0.39651
Comma—PRP—VBZ	0.0026170	0.0024638	0.0021702	0.0017973	M	0.29820
Comma—RB—Comma	0.0012810	0.0017110	0.0014582	0.0018165	F	0.45963
Comma—RB—IN	0.0017320	0.0015178	0.0022131	0.0018102	F	0.09427
Comma—VBG—DT	0.0012415	0.0012795	0.0014192	0.0014181	F	0.42872
OQuote—DT—NN	0.0046355	0.0030229	0.0049865	0.0037341	F	0.98246
OQuote—DT—NNP	0.0022130	0.0023225	0.0022014	0.0022112	M	0.94880

Findings *significant in this study* ($p < 0.05$) are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

Feature	Gender M Mean	Gender M Std Dev	Gender F Mean	Gender F Std Dev	Gender Prevalence	Mann-Whitney p-value
OQuote—IN—DT	0.0022550	0.0020446	0.0026014	0.0023091	F	0.32772
OQuote—JJ—NN	0.0023250	0.0020807	0.0017111	0.0015781	M	0.08154
OQuote—NN—VBD	0.0011137	0.0037956	0.0019762	0.0047608	F	0.17460
OQuote—NNP—CQuote	0.0013988	0.0020229	0.0016089	0.0018875	F	0.23895
OQuote—NNP—NNP	0.0034908	0.0048977	0.0042395	0.0058575	F	0.25057
OQuote—RB—Comma	0.0012157	0.0013547	0.0015234	0.0014920	F	0.13097
CC—DT—JJ	0.0014568	0.0014393	0.0012983	0.0014802	M	0.28147
CC—DT—NN	0.0054954	0.0031593	0.0051093	0.0024532	M	0.63893
CC—IN—DT	0.0028939	0.0019638	0.0027257	0.0020508	M	0.48903
CC—JJ—NN	0.0042492	0.0043488	0.0044206	0.0041048	F	0.58008
CC—JJ—NNS	0.0018352	0.0017895	0.0016493	0.0020195	M	0.25683
CC—NN—IN	0.0052564	0.0033374	0.0054743	0.0034267	F	0.59075
CC—NN—NN	0.0015891	0.0016799	0.0015298	0.0020433	M	0.34578
CC—NN—Period	0.0015554	0.0016762	0.0016530	0.0017226	F	0.70024
CC—NNP—Comma	0.0013244	0.0013362	0.0015371	0.0014387	F	0.42940
CC—NNP—NNP	0.0028133	0.0029203	0.0030188	0.0031452	F	0.76266
CC—NNP—VBD	0.0017206	0.0022599	0.0023906	0.0027842	F	0.09980
CC—NNS—IN	0.0023063	0.0032186	0.0022567	0.0021720	M	0.47364
CC—VBD—IN	0.0014032	0.0016838	0.0014455	0.0013997	F	0.67615
CD—CD—CD	0.0069626	0.0124964	0.0071070	0.0104130	F	0.71546
CD—CD—DT	0.0012986	0.0020737	0.0014422	0.0026282	F	0.88603
CD—CD—NN	0.0020584	0.0029718	0.0026445	0.0040035	F	0.29153
CD—Comma—CD	0.0058077	0.0037927	0.0060473	0.0041399	F	0.79200
CD—Comma—NNP	0.0034218	0.0029633	0.0028076	0.0026628	M	0.14312
CD—DT—NN	0.0011745	0.0017358	0.0013900	0.0026111	F	0.80495
CD—IN—DT	0.0026443	0.0028563	0.0024034	0.0020621	M	0.88875
CD—JJ—Period	0.0030851	0.0052507	0.0039232	0.0059952	F	0.31756
CD—NN—Colon	0.0017577	0.0026058	0.0021339	0.0034033	F	0.49883
CD—NNS—IN	0.0020323	0.0018176	0.0019070	0.0017782	M	0.59725
CQuote—IN—DT	0.0021476	0.0021668	0.0024428	0.0024759	F	0.39739
DT—IN—DT	0.0019761	0.0017324	0.0017961	0.0014681	M	0.65233
DT—JJ—CC	0.0018897	0.0016685	0.0016578	0.0016796	M	0.24904
DT—JJ—IN	0.0021692	0.0020639	0.0015058	0.0018606	M	0.01014
DT—JJ—JJ	0.0026763	0.0024553	0.0024238	0.0022872	M	0.35221

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

Feature	Gender M Mean	Gender M Std Dev	Gender F Mean	Gender F Std Dev	Gender Prevalence	Mann-Whitney p-value
DT—JJ—NN	0.0466386	0.0111081	0.0470346	0.0116428	F	0.58278
DT—JJ—NNP	0.0012231	0.0016722	0.0013576	0.0018806	F	0.81272
DT—JJ—NNS	0.0056278	0.0031912	0.0047240	0.0029729	M	0.03986
DT—JJ—Period	0.0015904	0.0027138	0.0012411	0.0019211	M	0.41184
DT—NN—CC	0.0105970	0.0056226	0.0109802	0.0054797	F	0.54524
DT—NN—Colon	0.0017425	0.0018323	0.0017730	0.0024871	F	0.51755
DT—NN—Comma	0.0180633	0.0068324	0.0188160	0.0067779	F	0.34605
DT—NN—DT	0.0018101	0.0015662	0.0015204	0.0014097	M	0.19359
DT—NN—IN	0.0875075	0.0183555	0.0883673	0.0200940	F	0.90022
DT—NN—JJ	0.0016034	0.0015055	0.0017720	0.0025380	F	0.58690
DT—NN—MD	0.0104112	0.0060332	0.0105084	0.0062309	F	0.93099
DT—NN—NN	0.0376355	0.0112952	0.0368693	0.0128861	M	0.24576
DT—NN—NNS	0.0045709	0.0041310	0.0044339	0.0030879	M	0.97834
DT—NN—OQuote	0.0021316	0.0025938	0.0025037	0.0029732	F	0.31737
DT—NN—Period	0.0223341	0.0078857	0.0211018	0.0077702	M	0.41918
DT—NN—POS	0.0062847	0.0077460	0.0062721	0.0057080	M	0.71158
DT—NN—PRP	0.0014825	0.0014848	0.0016078	0.0014702	F	0.36221
DT—NN—RB	0.0044152	0.0035685	0.0043998	0.0027357	M	0.62329
DT—NN—TO	0.0110497	0.0049158	0.0127692	0.0061335	F	0.08513
DT—NN—VBD	0.0183167	0.0081237	0.0175123	0.0073341	M	0.72904
DT—NN—VBG	0.0041100	0.0031556	0.0033196	0.0024697	M	0.07234
DT—NN—VBN	0.0096024	0.0104224	0.0101190	0.0099562	F	0.65372
DT—NN—VBZ	0.0203584	0.0081932	0.0180919	0.0073811	M	0.02809
DT—NN—WDT	0.0031348	0.0023258	0.0026591	0.0023343	M	0.10495
DT—NNP—CC	0.0012972	0.0015387	0.0013915	0.0018935	F	0.98618
DT—NNP—Comma	0.0018988	0.0022771	0.0015614	0.0025476	M	0.10184
DT—NNP—IN	0.0043706	0.0037281	0.0036825	0.0040241	M	0.10226
DT—NNP—MD	0.0043216	0.0057335	0.0036493	0.0041732	M	0.86276
DT—NNP—NN	0.0048090	0.0046385	0.0040008	0.0040387	M	0.18149
DT—NNP—NNP	0.0187257	0.0120501	0.0166466	0.0126333	M	0.18393
DT—NNP—Period	0.0023241	0.0032700	0.0019683	0.0028018	M	0.83519
DT—NNP—POS	0.0034759	0.0055596	0.0022794	0.0031299	M	0.08086
DT—NNP—VBD	0.0039376	0.0062802	0.0034367	0.0078975	M	0.01379
DT—NNP—VBZ	0.0037002	0.0042817	0.0030641	0.0031088	M	0.56390

Findings *significant in this study* ($p < 0.05$) are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
DT—NNS—CC	0.0032461	0.0058800	0.0024305	0.0031159	M	0.83478
DT—NNS—Comma	0.0029696	0.0020898	0.0030818	0.0020746	F	0.59229
DT—NNS—IN	0.0112807	0.0053627	0.0111249	0.0051753	M	0.93407
DT—NNS—Period	0.0028964	0.0022657	0.0029884	0.0023981	F	0.94638
DT—NNS—POS	0.0013644	0.0017851	0.0012620	0.0016975	M	0.75014
DT—NNS—RB	0.0016234	0.0015368	0.0018096	0.0017169	F	0.60838
DT—NNS—TO	0.0016618	0.0016851	0.0014954	0.0014891	M	0.64341
DT—NNS—VBD	0.0051480	0.0037476	0.0053979	0.0045422	F	0.96598
DT—NNS—VBP	0.0036539	0.0023814	0.0033380	0.0025037	M	0.29476
DT—OQuote—JJ	0.0016589	0.0026697	0.0012559	0.0016844	M	0.43334
DT—OQuote—NN	0.0024536	0.0042811	0.0031562	0.0056310	F	0.98220
DT—VBG—NN	0.0017857	0.0031961	0.0014812	0.0015354	M	0.80974
DT—VBN—NN	0.0112044	0.0105893	0.0115637	0.0102811	F	0.80486
EX—VBD—DT	0.0012846	0.0015055	0.0016520	0.0019928	F	0.46642
EX—VBZ—DT	0.0037959	0.0031170	0.0031714	0.0027461	M	0.16869
IN—CD—JJ	0.0023543	0.0039278	0.0029989	0.0043909	F	0.20537
IN—CD—NNS	0.0021174	0.0021759	0.0016474	0.0016351	M	0.25820
IN—DT—CD	0.0019301	0.0019698	0.0021409	0.0022721	F	0.70502
IN—DT—JJ	0.0277865	0.0081992	0.0257289	0.0075777	M	0.06702
IN—DT—NN	0.1245317	0.0262501	0.1258360	0.0297923	F	0.80700
IN—DT—NNP	0.0227326	0.0133644	0.0203307	0.0125281	M	0.18954
IN—DT—NNS	0.0205586	0.0071587	0.0198032	0.0084141	M	0.49168
IN—DT—OQuote	0.0019468	0.0027130	0.0021706	0.0033295	F	0.84682
IN—DT—VBG	0.0015773	0.0017556	0.0011308	0.0013016	M	0.12919
IN—DT—VBN	0.0078531	0.0076271	0.0078187	0.0068375	M	0.70547
IN—EX—VBZ	0.0019250	0.0019225	0.0015605	0.0015742	M	0.18683
IN—IN—DT	0.0040623	0.0029139	0.0034573	0.0021013	M	0.33619
IN—JJ—Comma	0.0015736	0.0018605	0.0012465	0.0014965	M	0.37038
IN—JJ—IN	0.0015286	0.0015745	0.0013517	0.0013409	M	0.69594
IN—JJ—NN	0.0118714	0.0066067	0.0114760	0.0073189	M	0.39147
IN—JJ—NNS	0.0067754	0.0036717	0.0062971	0.0035988	M	0.46600
IN—NN—CC	0.0043514	0.0028632	0.0041461	0.0026173	M	0.82205
IN—NN—Comma	0.0092124	0.0047380	0.0100870	0.0045856	F	0.14061
IN—NN—CQuote	0.0027301	0.0071447	0.0037857	0.0072446	F	0.13015

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features… (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
IN—NN—DT	0.0044446	0.0029477	0.0041297	0.0029505	M	0.42139
IN—NN—IN	0.0177406	0.0064139	0.0186106	0.0059364	F	0.15063
IN—NN—NN	0.0126313	0.0069427	0.0136129	0.0078040	F	0.51810
IN—NN—NNS	0.0053711	0.0029554	0.0057114	0.0032292	F	0.64448
IN—NN—Period	0.0121107	0.0052763	0.0130653	0.0047711	F	0.15887
IN—NN—TO	0.0054484	0.0030178	0.0061418	0.0033389	F	0.10644
IN—NN—VBZ	0.0023866	0.0021368	0.0021772	0.0018637	M	0.54317
IN—NNP—CC	0.0035372	0.0027218	0.0044726	0.0037439	F	0.19910
IN—NNP—CD	0.0080487	0.0061367	0.0089181	0.0060291	F	0.21275
IN—NNP—Comma	0.0052262	0.0044800	0.0050289	0.0030721	M	0.57656
IN—NNP—IN	0.0031538	0.0027001	0.0033054	0.0026329	F	0.57146
IN—NNP—NN	0.0012292	0.0017897	0.0017754	0.0021526	F	0.01448
IN—NNP—NNP	0.0136071	0.0107103	0.0139938	0.0128879	F	0.82004
IN—NNP—Period	0.0040272	0.0029138	0.0038758	0.0029098	M	0.82706
IN—NNP—POS	0.0054397	0.0054121	0.0059961	0.0054294	F	0.29080
IN—NNP—VBD	0.0054578	0.0041630	0.0064661	0.0052116	F	0.14024
IN—NNP—VBZ	0.0017437	0.0018136	0.0021041	0.0020981	F	0.30863
IN—NNS—IN	0.0038343	0.0026397	0.0037382	0.0027670	M	0.63241
IN—PRP—MD	0.0020053	0.0015495	0.0021563	0.0018419	F	0.83665
IN—PRP—VBD	0.0065388	0.0062667	0.0067909	0.0056157	F	0.66125
IN—PRP—VBP	0.0014128	0.0012878	0.0014336	0.0014754	F	0.78849
IN—PRP—VBZ	0.0030905	0.0026490	0.0028150	0.0023054	M	0.59200
IN—PRP$—JJ	0.0023150	0.0024902	0.0024524	0.0026303	F	0.88738
IN—PRP$—NN	0.0095092	0.0056763	0.0095210	0.0064903	F	0.60064
IN—PRP$—NNS	0.0023216	0.0028650	0.0017041	0.0021482	M	0.23587
IN—VBG—DT	0.0042007	0.0028718	0.0035900	0.0023042	M	0.10421
IN—VBN—IN	0.0018018	0.0020142	0.0017451	0.0018443	M	0.85772
JJ—CC—JJ	0.0031962	0.0026381	0.0033436	0.0033558	F	0.65156
JJ—Comma—CC	0.0015204	0.0018146	0.0014843	0.0018007	M	0.78450
JJ—Comma—DT	0.0016819	0.0015440	0.0015445	0.0013977	M	0.68685
JJ—DT—NN	0.0027053	0.0019463	0.0022426	0.0019189	M	0.08344
JJ—IN—DT	0.0091625	0.0049531	0.0086008	0.0039139	M	0.65192
JJ—IN—NN	0.0029486	0.0021672	0.0029226	0.0032265	M	0.16183
JJ—IN—NNP	0.0021285	0.0019649	0.0021896	0.0022890	F	0.63163

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
JJ—JJ—NN	0.0030593	0.0026195	0.0033705	0.0025040	F	0.23058
JJ—JJ—NNS	0.0014271	0.0015079	0.0014788	0.0013713	F	0.60080
JJ—NN—CC	0.0045641	0.0025680	0.0045432	0.0023422	M	0.93818
JJ—NN—Comma	0.0089914	0.0044965	0.0093977	0.0042001	F	0.63524
JJ—NN—IN	0.0255067	0.0068551	0.0269574	0.0075267	F	0.30172
JJ—NN—MD	0.0032064	0.0023018	0.0030223	0.0024846	M	0.39996
JJ—NN—NN	0.0078031	0.0044258	0.0080380	0.0038105	F	0.38648
JJ—NN—NNS	0.0042499	0.0029798	0.0040149	0.0026448	M	0.83514
JJ—NN—Period	0.0125515	0.0060360	0.0118411	0.0065202	M	0.26687
JJ—NN—TO	0.0039115	0.0022782	0.0042528	0.0030081	F	0.84930
JJ—NN—VBD	0.0029808	0.0023674	0.0026650	0.0021586	M	0.33193
JJ—NN—VBZ	0.0043064	0.0026708	0.0040178	0.0025352	M	0.58185
JJ—NN—WDT	0.0022068	0.0018659	0.0020606	0.0018438	M	0.47037
JJ—NNS—CC	0.0017551	0.0015685	0.0019456	0.0018477	F	0.49418
JJ—NNS—Comma	0.0028703	0.0019468	0.0030664	0.0027212	F	0.85322
JJ—NNS—IN	0.0078710	0.0034148	0.0073370	0.0037368	M	0.24158
JJ—NNS—Period	0.0040987	0.0024266	0.0039927	0.0026266	M	0.42291
JJ—NNS—TO	0.0015644	0.0013984	0.0014274	0.0012975	M	0.65127
JJ—NNS—VBP	0.0015722	0.0016961	0.0015268	0.0014927	M	0.86379
JJ—TO—DT	0.0036822	0.0026202	0.0034924	0.0025305	M	0.67533
JJ—TO—VB	0.0045324	0.0030617	0.0047271	0.0042177	F	0.83110
MD—RB—VB	0.0103626	0.0048762	0.0093274	0.0051304	M	0.04639
MD—VB—DT	0.0089800	0.0048635	0.0083141	0.0045729	M	0.33041
MD—VB—IN	0.0048024	0.0028773	0.0050421	0.0028882	F	0.49494
MD—VB—JJ	0.0033667	0.0025016	0.0029303	0.0023337	M	0.16217
MD—VB—NN	0.0015583	0.0017705	0.0014432	0.0012491	M	0.67614
MD—VB—RB	0.0020517	0.0017217	0.0019994	0.0015867	M	0.91733
MD—VB—TO	0.0020108	0.0019454	0.0015855	0.0013502	M	0.31735
MD—VB—VBN	0.0155561	0.0068861	0.0153923	0.0073537	M	0.89409
NN—CC—DT	0.0032195	0.0020924	0.0033399	0.0022261	F	0.78999
NN—CC—IN	0.0016289	0.0013676	0.0020039	0.0019532	F	0.29541
NN—CC—JJ	0.0030917	0.0039850	0.0033240	0.0034899	F	0.23148
NN—CC—NN	0.0102687	0.0057581	0.0101151	0.0050761	M	0.99278
NN—CC—RB	0.0018759	0.0019482	0.0016117	0.0014964	M	0.72799

Findings *significant in this study* ($p < 0.05$) are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features. . . (continued)

Feature	Gender M Mean	Gender M Std Dev	Gender F Mean	Gender F Std Dev	Gender Prevalence	Mann-Whitney p-value
NN—CC—VBD	0.0025311	0.0022169	0.0031585	0.0029393	F	0.16590
NN—CD—CD	0.0022616	0.0040869	0.0025893	0.0036462	F	0.18247
NN—Colon—Colon	0.0011199	0.0014813	0.0015573	0.0022250	F	0.15802
NN—Comma—CC	0.0090913	0.0050140	0.0096409	0.0045035	F	0.37177
NN—Comma—DT	0.0099279	0.0047723	0.0103200	0.0045849	F	0.48519
NN—Comma—IN	0.0058803	0.0033815	0.0054300	0.0040344	M	0.16389
NN—Comma—JJ	0.0022940	0.0016142	0.0023513	0.0019109	F	0.90521
NN—Comma—NN	0.0023880	0.0020022	0.0028905	0.0025899	F	0.33113
NN—Comma—NNP	0.0058849	0.0037015	0.0078224	0.0044691	F	0.00100
NN—Comma—PRP	0.0035883	0.0024842	0.0033055	0.0024858	M	0.29861
NN—Comma—RB	0.0033252	0.0019325	0.0033200	0.0022007	M	0.83816
NN—Comma—VBD	0.0014103	0.0012103	0.0018817	0.0015741	F	0.06427
NN—Comma—VBG	0.0020276	0.0017443	0.0023387	0.0021702	F	0.50207
NN—Comma—WDT	0.0015229	0.0015746	0.0013944	0.0015525	M	0.38827
NN—CQuote—IN	0.0016934	0.0018365	0.0022941	0.0027368	F	0.14417
NN—CQuote—NN	0.0014410	0.0045433	0.0016170	0.0039696	F	0.45123
NN—DT—JJ	0.0017336	0.0015386	0.0015083	0.0014058	M	0.36095
NN—DT—NN	0.0094834	0.0042277	0.0091286	0.0040934	M	0.53073
NN—IN—CD	0.0029623	0.0030233	0.0034509	0.0033316	F	0.35542
NN—IN—DT	0.0849747	0.0202006	0.0840127	0.0238096	M	0.79302
NN—IN—IN	0.0021740	0.0018499	0.0019098	0.0017398	M	0.31823
NN—IN—JJ	0.0103536	0.0058279	0.0105053	0.0073170	F	0.57306
NN—IN—NN	0.0387257	0.0135754	0.0395190	0.0132146	F	0.65845
NN—IN—NNP	0.0176645	0.0083192	0.0210109	0.0093638	F	0.00927
NN—IN—NNS	0.0069821	0.0034546	0.0064713	0.0036319	M	0.21228
NN—IN—OQuote	0.0019812	0.0024991	0.0020597	0.0024891	F	0.74106
NN—IN—PRP	0.0042012	0.0030582	0.0042614	0.0027082	F	0.73000
NN—IN—PRP$	0.0056790	0.0042235	0.0058548	0.0043833	F	0.91456
NN—IN—RB	0.0014154	0.0013041	0.0015552	0.0014168	F	0.57451
NN—IN—TO	0.0017004	0.0022281	0.0014534	0.0016572	M	0.90690
NN—IN—VBG	0.0046088	0.0032103	0.0043946	0.0023030	M	0.69819
NN—IN—VBN	0.0014964	0.0016012	0.0014202	0.0013299	M	0.79138
NN—IN—WDT	0.0018203	0.0018397	0.0018521	0.0018603	F	0.87092
NN—JJ—NN	0.0014943	0.0017811	0.0012276	0.0013928	M	0.61728

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
NN—MD—RB	0.0035974	0.0023079	0.0034202	0.0023158	M	0.54944
NN—MD—VB	0.0179763	0.0075608	0.0178112	0.0081321	M	0.71838
NN—NN—CC	0.0041348	0.0031869	0.0044768	0.0026894	F	0.19750
NN—NN—Comma	0.0081594	0.0041833	0.0082751	0.0043255	F	0.93818
NN—NN—DT	0.0018150	0.0016630	0.0017064	0.0016620	M	0.63725
NN—NN—IN	0.0207949	0.0065482	0.0209532	0.0065334	F	0.85437
NN—NN—MD	0.0025596	0.0021535	0.0023043	0.0020669	M	0.34395
NN—NN—NN	0.0068097	0.0044296	0.0058240	0.0048085	M	0.04203
NN—NN—NNS	0.0025619	0.0022154	0.0022976	0.0020338	M	0.53933
NN—NN—Period	0.0118967	0.0054075	0.0115759	0.0057620	M	0.63156
NN—NN—TO	0.0028175	0.0020778	0.0033222	0.0022303	F	0.13484
NN—NN—VBD	0.0035353	0.0028280	0.0032656	0.0023228	M	0.86847
NN—NN—VBZ	0.0042576	0.0026005	0.0037121	0.0025702	M	0.14832
NN—NN—WDT	0.0017133	0.0019598	0.0015427	0.0017101	M	0.67997
NN—NNP—NNP	0.0019239	0.0018697	0.0014707	0.0014604	M	0.11029
NN—NNS—CC	0.0015566	0.0026832	0.0014735	0.0016429	M	0.69467
NN—NNS—Comma	0.0020667	0.0017962	0.0026079	0.0019940	F	0.05146
NN—NNS—DT	0.0016805	0.0018437	0.0015885	0.0016164	M	0.96212
NN—NNS—IN	0.0055038	0.0027930	0.0056072	0.0029826	F	0.80100
NN—NNS—Period	0.0037471	0.0026881	0.0032416	0.0026873	M	0.14370
NN—POS—NN	0.0052083	0.0049263	0.0060158	0.0052918	F	0.23321
NN—RB—TO	0.0015807	0.0017590	0.0017695	0.0026639	F	0.87371
NN—RB—VBZ	0.0021898	0.0019811	0.0017739	0.0015086	M	0.15433
NN—TO—DT	0.0059670	0.0031608	0.0065890	0.0035776	F	0.34276
NN—TO—NN	0.0021507	0.0018307	0.0021054	0.0019678	M	0.63465
NN—TO—NNP	0.0017136	0.0023188	0.0017127	0.0021362	M	0.83636
NN—TO—NNS	0.0020620	0.0025885	0.0018946	0.0023787	M	0.67760
NN—TO—VB	0.0166442	0.0063993	0.0204396	0.0081413	F	0.00097
NN—VBD—DT	0.0039523	0.0032946	0.0037630	0.0028922	M	0.90733
NN—VBD—IN	0.0084571	0.0052755	0.0089302	0.0059660	F	0.82910
NN—VBD—JJ	0.0021292	0.0022794	0.0018871	0.0021082	M	0.45918
NN—VBD—RB	0.0046435	0.0036288	0.0042518	0.0030222	M	0.65184
NN—VBD—TO	0.0018583	0.0014525	0.0016091	0.0016779	M	0.09537
NN—VBD—VBN	0.0073903	0.0039580	0.0066745	0.0041007	M	0.19041

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

Feature	Gender M Mean	Gender M Std Dev	Gender F Mean	Gender F Std Dev	Gender Prevalence	Mann-Whitney p-value
NN—VBG—DT	0.0022365	0.0018018	0.0023131	0.0020782	F	0.98345
NN—VBG—IN	0.0013569	0.0013025	0.0014296	0.0014898	F	0.99158
NN—VBN—IN	0.0102993	0.0115204	0.0117528	0.0110862	F	0.29028
NN—VBZ—DT	0.0055179	0.0029749	0.0053432	0.0035534	M	0.43955
NN—VBZ—IN	0.0052461	0.0033614	0.0042501	0.0026341	M	0.03421
NN—VBZ—JJ	0.0049067	0.0032989	0.0045295	0.0027908	M	0.64076
NN—VBZ—RB	0.0070406	0.0034740	0.0062889	0.0032205	M	0.12058
NN—VBZ—TO	0.0015516	0.0015238	0.0016488	0.0017611	F	0.87062
NN—VBZ—VBN	0.0079613	0.0051130	0.0070256	0.0040039	M	0.33041
NN—WDT—DT	0.0017991	0.0016346	0.0018718	0.0020202	F	0.69614
NN—WDT—VBZ	0.0019816	0.0019163	0.0017311	0.0015466	M	0.48628
NNP—CC—NNP	0.0074148	0.0057552	0.0087325	0.0064322	F	0.13410
NNP—CD—CD	0.0033679	0.0063530	0.0040110	0.0057661	F	0.30416
NNP—CD—Comma	0.0062599	0.0040172	0.0063729	0.0040967	F	0.93406
NNP—Comma—CC	0.0015243	0.0018003	0.0014732	0.0016624	M	0.93546
NNP—Comma—DT	0.0058160	0.0039099	0.0065157	0.0034707	F	0.13144
NNP—Comma—IN	0.0021672	0.0020188	0.0022735	0.0018968	F	0.49786
NNP—Comma—NNP	0.0051721	0.0041905	0.0068021	0.0052635	F	0.00611
NNP—CQuote—Colon	0.0013628	0.0019910	0.0015915	0.0019377	F	0.29693
NNP—IN—CD	0.0033850	0.0040504	0.0030888	0.0033698	M	0.89104
NNP—IN—DT	0.0048817	0.0029962	0.0049174	0.0033298	F	0.90227
NNP—IN—NN	0.0014203	0.0014408	0.0013482	0.0013402	M	0.93714
NNP—IN—NNP	0.0073079	0.0057697	0.0067565	0.0061297	M	0.34605
NNP—MD—RB	0.0019285	0.0020977	0.0018240	0.0020156	M	0.87964
NNP—MD—VB	0.0071290	0.0055990	0.0067226	0.0045226	M	0.91251
NNP—NN—IN	0.0018484	0.0017059	0.0016956	0.0016676	M	0.49714
NNP—NNP—CC	0.0030101	0.0034538	0.0030610	0.0039498	F	0.93577
NNP—NNP—Comma	0.0079365	0.0050446	0.0094484	0.0057158	F	0.04541
NNP—NNP—CQuote	0.0025609	0.0037702	0.0023407	0.0033452	M	0.89154
NNP—NNP—IN	0.0092664	0.0062090	0.0079195	0.0057855	M	0.12496
NNP—NNP—NN	0.0022822	0.0025632	0.0019976	0.0026597	M	0.21397
NNP—NNP—NNP	0.0118157	0.0125323	0.0119229	0.0106246	F	0.47314
NNP—NNP—OQuote	0.0011612	0.0015717	0.0015277	0.0018199	F	0.22082
NNP—NNP—Period	0.0067893	0.0050725	0.0077395	0.0047908	F	0.05347

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

	Gender M		Gender F		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
NNP—NNP—POS	0.0055584	0.0061293	0.0056367	0.0074407	F	0.57606
NNP—NNP—RB	0.0019246	0.0019316	0.0020410	0.0024484	F	0.78130
NNP—NNP—VBD	0.0113353	0.0103402	0.0113315	0.0123032	M	0.82306
NNP—NNP—VBZ	0.0060286	0.0041204	0.0048980	0.0036977	M	0.04228
NNP—POS—JJ	0.0023781	0.0026749	0.0029201	0.0033059	F	0.27664
NNP—POS—NN	0.0135046	0.0078595	0.0149028	0.0094996	F	0.45267
NNP—POS—NNP	0.0015289	0.0022574	0.0018289	0.0026274	F	0.50464
NNP—POS—NNS	0.0027977	0.0035687	0.0025675	0.0028670	M	0.77523
NNP—RB—VBD	0.0033709	0.0023976	0.0037413	0.0028103	F	0.53062
NNP—RB—VBZ	0.0014969	0.0013639	0.0015083	0.0015158	F	0.80540
NNP—TO—VB	0.0025544	0.0021852	0.0031398	0.0022159	F	0.03641
NNP—VBD—DT	0.0084057	0.0039088	0.0086062	0.0044520	F	0.94950
NNP—VBD—IN	0.0061100	0.0036538	0.0063332	0.0036822	F	0.65285
NNP—VBD—NNP	0.0020125	0.0018230	0.0030494	0.0032193	F	0.05538
NNP—VBD—RB	0.0040955	0.0025323	0.0067504	0.0091776	F	0.05112
NNP—VBD—TO	0.0028764	0.0024726	0.0035553	0.0021121	F	0.00844
NNP—VBD—VBN	0.0029860	0.0026793	0.0028997	0.0027463	M	0.66666
NNP—VBZ—DT	0.0038688	0.0022799	0.0038387	0.0027449	M	0.60693
NNP—VBZ—IN	0.0029830	0.0023349	0.0022922	0.0021272	M	0.03024
NNP—VBZ—RB	0.0031329	0.0031521	0.0028932	0.0025116	M	0.74342
NNP—VBZ—VBN	0.0036850	0.0030372	0.0032683	0.0026473	M	0.42459
NNS—CC—NNS	0.0044543	0.0072719	0.0038362	0.0038409	M	0.82401
NNS—Comma—CC	0.0033502	0.0024239	0.0031757	0.0023794	M	0.60596
NNS—Comma—DT	0.0019699	0.0016047	0.0023841	0.0017472	F	0.09366
NNS—Comma—IN	0.0016561	0.0014466	0.0013630	0.0013898	M	0.12697
NNS—Comma—NNS	0.0016091	0.0023424	0.0017897	0.0023675	F	0.62827
NNS—DT—NN	0.0035451	0.0024926	0.0034389	0.0024701	M	0.71439
NNS—IN—DT	0.0170475	0.0063741	0.0158188	0.0055864	M	0.23489
NNS—IN—JJ	0.0037217	0.0029118	0.0033676	0.0025151	M	0.50205
NNS—IN—NN	0.0082167	0.0042742	0.0088416	0.0055817	F	0.79701
NNS—IN—NNP	0.0035078	0.0026391	0.0036382	0.0029585	F	0.87978
NNS—IN—NNS	0.0024865	0.0021599	0.0024236	0.0022806	M	0.61181
NNS—IN—PRP$	0.0015379	0.0017437	0.0013097	0.0016025	M	0.37529
NNS—IN—VBG	0.0023138	0.0017057	0.0022231	0.0019432	M	0.32068

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
NNS—MD—VB	0.0044136	0.0031538	0.0036753	0.0024876	M	0.14760
NNS—POS—NN	0.0019712	0.0027964	0.0018283	0.0024006	M	0.81806
NNS—TO—DT	0.0018574	0.0019352	0.0017790	0.0016964	M	0.96462
NNS—TO—VB	0.0046419	0.0028006	0.0044953	0.0028774	M	0.66407
NNS—VBD—DT	0.0014870	0.0013824	0.0012866	0.0013468	M	0.26247
NNS—VBD—IN	0.0028862	0.0023051	0.0033286	0.0027796	F	0.24135
NNS—VBD—RB	0.0014690	0.0018935	0.0014346	0.0016029	M	0.96636
NNS—VBD—VBN	0.0018252	0.0019746	0.0018948	0.0018335	F	0.65532
NNS—VBN—IN	0.0028809	0.0027425	0.0031766	0.0027551	F	0.34598
NNS—VBP—IN	0.0020454	0.0017857	0.0017996	0.0015271	M	0.46455
NNS—VBP—RB	0.0021284	0.0019238	0.0024280	0.0019711	F	0.26157
NNS—VBP—VBN	0.0023831	0.0022482	0.0025071	0.0021043	F	0.46005
NNS—WDT—DT	0.0012680	0.0012721	0.0013931	0.0015288	F	0.89922
POS—JJ—NN	0.0024699	0.0021648	0.0028126	0.0026298	F	0.61299
POS—NN—Comma	0.0014146	0.0014316	0.0020867	0.0021449	F	0.03377
POS—NN—IN	0.0064137	0.0040642	0.0064458	0.0046834	F	0.76822
POS—NN—NN	0.0020332	0.0017667	0.0022657	0.0026642	F	0.72495
POS—NN—Period	0.0027828	0.0021440	0.0031955	0.0028122	F	0.56027
POS—NN—TO	0.0018352	0.0018403	0.0025637	0.0023612	F	0.03483
POS—NN—VBD	0.0017419	0.0028448	0.0015080	0.0020862	M	0.80846
POS—NN—VBZ	0.0014953	0.0016817	0.0012672	0.0016331	M	0.25861
PRP—IN—DT	0.0020183	0.0016137	0.0016131	0.0014883	M	0.05137
PRP—MD—RB	0.0016189	0.0016978	0.0014984	0.0017423	M	0.38709
PRP—MD—VB	0.0054460	0.0030104	0.0050814	0.0030533	M	0.22833
PRP—VBD—DT	0.0023733	0.0023632	0.0028216	0.0027057	F	0.12860
PRP—VBD—IN	0.0021713	0.0023038	0.0022033	0.0021667	F	0.75974
PRP—VBD—RB	0.0034328	0.0034462	0.0032539	0.0030854	M	1.00000
PRP—VBD—TO	0.0012603	0.0019465	0.0015707	0.0018707	F	0.07339
PRP—VBD—VBN	0.0020080	0.0020822	0.0024711	0.0025119	F	0.27918
PRP—VBZ—DT	0.0017775	0.0015165	0.0013411	0.0013657	M	0.03600
PRP—VBZ—JJ	0.0022582	0.0020142	0.0025056	0.0020191	F	0.27847
PRP—VBZ—RB	0.0028782	0.0022934	0.0023914	0.0022252	M	0.07440
PRP—VBZ—VBN	0.0017280	0.0020406	0.0012173	0.0012731	M	0.23253
PRP$—JJ—NN	0.0025908	0.0023881	0.0025091	0.0023790	M	0.78458

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

	Gender M		Gender F		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
PRP$—NN—Comma	0.0020348	0.0020208	0.0023657	0.0022378	F	0.27404
PRP$—NN—IN	0.0045188	0.0029304	0.0046516	0.0028702	F	0.52904
PRP$—NN—NN	0.0020994	0.0019036	0.0016817	0.0016507	M	0.10478
PRP$—NN—Period	0.0034184	0.0025351	0.0034722	0.0028600	F	0.79096
RB—Comma—DT	0.0049591	0.0034099	0.0051101	0.0029577	F	0.48842
RB—Comma—IN	0.0020464	0.0022290	0.0020136	0.0017330	M	0.76507
RB—Comma—NNP	0.0020866	0.0019674	0.0027337	0.0033241	F	0.33771
RB—DT—JJ	0.0016140	0.0015933	0.0014198	0.0014686	M	0.45872
RB—DT—NN	0.0047540	0.0030972	0.0037187	0.0025519	M	0.02594
RB—IN—DT	0.0059860	0.0028267	0.0062479	0.0027781	F	0.47155
RB—IN—NN	0.0018490	0.0016403	0.0014544	0.0012277	M	0.14311
RB—JJ—IN	0.0037502	0.0033617	0.0031299	0.0024503	M	0.47697
RB—JJ—NN	0.0018357	0.0020276	0.0017466	0.0015424	M	0.51948
RB—JJ—Period	0.0014487	0.0016453	0.0015987	0.0016807	F	0.60819
RB—JJ—TO	0.0017300	0.0018197	0.0017269	0.0018825	M	0.85063
RB—TO—DT	0.0023695	0.0030990	0.0028326	0.0039228	F	0.37633
RB—TO—VB	0.0024412	0.0023198	0.0019534	0.0018406	M	0.08720
RB—VB—DT	0.0071079	0.0035261	0.0078318	0.0046260	F	0.45735
RB—VB—IN	0.0033962	0.0022367	0.0033948	0.0029697	M	0.34528
RB—VB—TO	0.0014134	0.0015786	0.0014818	0.0013962	F	0.58592
RB—VB—VBN	0.0040230	0.0023729	0.0044371	0.0032402	F	0.95361
RB—VBD—DT	0.0022571	0.0022741	0.0020866	0.0017891	M	0.92557
RB—VBD—IN	0.0021754	0.0019914	0.0021261	0.0017884	M	0.86827
RB—VBN—CC	0.0013732	0.0016563	0.0014301	0.0015776	F	0.67512
RB—VBN—DT	0.0018887	0.0019186	0.0019238	0.0017757	F	0.65251
RB—VBN—IN	0.0060039	0.0028779	0.0056128	0.0026419	M	0.47155
RB—VBN—TO	0.0015917	0.0016609	0.0017232	0.0018830	F	0.59142
RB—VBZ—DT	0.0022125	0.0018494	0.0019438	0.0015454	M	0.40744
RB—VBZ—IN	0.0021460	0.0019338	0.0019714	0.0016067	M	0.73373
TO—DT—JJ	0.0029323	0.0020035	0.0034112	0.0022728	F	0.06950
TO—DT—NN	0.0135566	0.0050716	0.0142557	0.0061324	F	0.62605
TO—DT—NNP	0.0018620	0.0018620	0.0015523	0.0025478	M	0.02148
TO—DT—NNS	0.0019934	0.0020761	0.0016557	0.0017304	M	0.20590
TO—JJ—NN	0.0011120	0.0018824	0.0016480	0.0023384	F	0.02030

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

Feature	Gender M Mean	Gender M Std Dev	Gender F Mean	Gender F Std Dev	Gender Prevalence	Mann-Whitney p-value
TO—NN—DT	0.0014880	0.0014572	0.0015139	0.0017246	F	0.70322
TO—NN—IN	0.0023605	0.0024527	0.0022311	0.0024151	M	0.49090
TO—NNP—NNP	0.0016387	0.0025956	0.0015833	0.0022972	M	0.81079
TO—VB—DT	0.0186352	0.0067570	0.0201228	0.0076144	F	0.24210
TO—VB—IN	0.0070343	0.0036950	0.0082141	0.0045387	F	0.12213
TO—VB—JJ	0.0040926	0.0028000	0.0042014	0.0029121	F	0.77117
TO—VB—NN	0.0047010	0.0026314	0.0050483	0.0027896	F	0.26575
TO—VB—NNP	0.0016089	0.0019177	0.0021957	0.0027262	F	0.22295
TO—VB—NNS	0.0024879	0.0019586	0.0026253	0.0019757	F	0.62678
TO—VB—Period	0.0014700	0.0027536	0.0014327	0.0017408	M	0.81856
TO—VB—PRP$	0.0024764	0.0024267	0.0024097	0.0018837	M	0.60982
TO—VB—RB	0.0015143	0.0014117	0.0015095	0.0015496	M	0.68440
TO—VB—VBN	0.0050925	0.0038889	0.0047807	0.0042483	M	0.32325
VB—DT—JJ	0.0060790	0.0037097	0.0062331	0.0033493	F	0.63524
VB—DT—NN	0.0218589	0.0078356	0.0234141	0.0077582	F	0.12980
VB—DT—NNP	0.0019354	0.0023163	0.0016768	0.0021311	M	0.62361
VB—DT—NNS	0.0029084	0.0024377	0.0029644	0.0024566	F	0.79078
VB—DT—VBN	0.0015576	0.0019459	0.0015206	0.0019831	M	0.68708
VB—IN—DT	0.0078398	0.0041380	0.0082047	0.0038796	F	0.40227
VB—IN—NN	0.0023643	0.0018295	0.0024800	0.0020230	F	0.78975
VB—IN—NNP	0.0010833	0.0014275	0.0018256	0.0021117	F	0.01163
VB—JJ—NN	0.0024159	0.0021428	0.0022801	0.0022897	M	0.43972
VB—JJ—NNS	0.0018132	0.0016524	0.0017804	0.0014644	M	0.87795
VB—NN—IN	0.0027227	0.0020354	0.0029449	0.0019858	F	0.29917
VB—NNS—IN	0.0014154	0.0014434	0.0015534	0.0016699	F	0.66266
VB—PRP$—NN	0.0021763	0.0018655	0.0021731	0.0019650	M	0.78842
VB—TO—VB	0.0023514	0.0021075	0.0020639	0.0017124	M	0.55382
VB—VBN—DT	0.0030413	0.0024624	0.0029010	0.0028769	M	0.32186
VB—VBN—IN	0.0107965	0.0058787	0.0108327	0.0054746	F	0.82809
VB—VBN—Period	0.0027985	0.0023074	0.0026953	0.0029257	M	0.28749
VB—VBN—RB	0.0015194	0.0020225	0.0018207	0.0021549	F	0.17461
VB—VBN—TO	0.0022212	0.0021302	0.0017105	0.0015060	M	0.20172
VBD—DT—JJ	0.0043949	0.0027859	0.0048950	0.0026467	F	0.14952
VBD—DT—NN	0.0137500	0.0074535	0.0138073	0.0057685	F	0.36080

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

Feature	Gender M Mean	Gender M Std Dev	Gender F Mean	Gender F Std Dev	Gender Prevalence	Mann-Whitney p-value
VBD—IN—DT	0.0103896	0.0044572	0.0106128	0.0048764	F	0.44647
VBD—IN—NN	0.0038872	0.0045186	0.0047482	0.0050998	F	0.24507
VBD—IN—NNP	0.0040229	0.0027568	0.0041810	0.0029924	F	0.84320
VBD—JJ—IN	0.0021521	0.0024331	0.0019209	0.0021031	M	0.61035
VBD—NN—IN	0.0017630	0.0016493	0.0018272	0.0014897	F	0.46352
VBD—NNP—NNP	0.0015481	0.0017840	0.0018149	0.0023061	F	0.65913
VBD—RB—JJ	0.0024874	0.0028578	0.0020765	0.0023016	M	0.45290
VBD—RB—VB	0.0046468	0.0037337	0.0065841	0.0071643	F	0.04428
VBD—RB—VBN	0.0046006	0.0028444	0.0050299	0.0030596	F	0.42662
VBD—TO—VB	0.0051959	0.0040723	0.0066844	0.0037269	F	0.00074
VBD—VBN—IN	0.0067288	0.0040726	0.0072122	0.0044893	F	0.46757
VBD—VBN—Period	0.0013726	0.0013302	0.0016231	0.0014922	F	0.26910
VBD—VBN—TO	0.0018673	0.0015955	0.0015833	0.0015524	M	0.18132
VBG—DT—JJ	0.0025663	0.0021880	0.0023740	0.0018899	M	0.52387
VBG—DT—NN	0.0098848	0.0046124	0.0092102	0.0045982	M	0.20388
VBG—IN—DT	0.0040590	0.0033581	0.0039983	0.0029571	M	0.76919
VBG—NN—IN	0.0022235	0.0020053	0.0021323	0.0019049	M	0.89350
VBN—CC—VBN	0.0020469	0.0019431	0.0023013	0.0021834	F	0.44117
VBN—DT—JJ	0.0017824	0.0017138	0.0017511	0.0015122	M	0.83912
VBN—DT—NN	0.0079268	0.0038681	0.0072220	0.0036438	M	0.16785
VBN—IN—DT	0.0216821	0.0073754	0.0219085	0.0076196	F	0.73780
VBN—IN—JJ	0.0033238	0.0025232	0.0026698	0.0019384	M	0.15864
VBN—IN—NN	0.0135124	0.0125375	0.0160251	0.0136824	F	0.21899
VBN—IN—NNP	0.0045387	0.0031958	0.0051475	0.0030979	F	0.08965
VBN—IN—PRP	0.0017802	0.0015524	0.0018263	0.0016632	F	0.94096
VBN—NN—IN	0.0043864	0.0037915	0.0040908	0.0033244	M	0.82692
VBN—NN—MD	0.0012766	0.0019069	0.0014580	0.0021522	F	0.81685
VBN—NN—NN	0.0016203	0.0018027	0.0012175	0.0015684	M	0.07994
VBN—NN—Period	0.0021074	0.0020770	0.0022018	0.0019514	F	0.60250
VBN—NN—VBN	0.0018344	0.0023399	0.0019105	0.0022878	F	0.75693
VBN—NNS—IN	0.0015525	0.0014258	0.0014042	0.0012720	M	0.52734
VBN—RB—TO	0.0011623	0.0021670	0.0014716	0.0024855	F	0.13791
VBN—TO—DT	0.0021962	0.0024346	0.0017241	0.0017912	M	0.41086
VBN—TO—NN	0.0017324	0.0019978	0.0015240	0.0017659	M	0.59871

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.5: All POS trigram features... (continued)

Feature	Gender M Mean	Gender M Std Dev	Gender F Mean	Gender F Std Dev	Gender Prevalence	Mann-Whitney p-value
VBN—TO—VB	0.0053529	0.0034322	0.0052234	0.0030344	M	0.94744
VBN—VBN—IN	0.0016599	0.0016273	0.0016821	0.0027016	F	0.27599
VBP—DT—NN	0.0016750	0.0013780	0.0016307	0.0014886	M	0.62738
VBP—IN—DT	0.0023118	0.0018019	0.0023408	0.0017490	F	0.94010
VBP—RB—VBN	0.0014429	0.0015779	0.0016336	0.0015650	F	0.28538
VBP—VBN—IN	0.0017448	0.0015402	0.0016511	0.0015298	M	0.68429
VBZ—DT—JJ	0.0044779	0.0025430	0.0042342	0.0023519	M	0.51893
VBZ—DT—NN	0.0137030	0.0047805	0.0126029	0.0053347	M	0.04403
VBZ—IN—DT	0.0065411	0.0036318	0.0053767	0.0028089	M	0.04126
VBZ—IN—NN	0.0013868	0.0012219	0.0014175	0.0014765	F	0.80540
VBZ—JJ—IN	0.0036088	0.0025122	0.0036848	0.0025462	F	0.72802
VBZ—JJ—TO	0.0018057	0.0016295	0.0021320	0.0017850	F	0.17958
VBZ—RB—DT	0.0019875	0.0019014	0.0017050	0.0015657	M	0.51459
VBZ—RB—JJ	0.0032664	0.0022825	0.0032111	0.0023770	M	0.89608
VBZ—RB—RB	0.0018327	0.0017087	0.0017056	0.0017250	M	0.55999
VBZ—RB—VB	0.0037387	0.0026434	0.0039605	0.0028789	F	0.74658
VBZ—RB—VBN	0.0040316	0.0029245	0.0033266	0.0023358	M	0.10836
VBZ—TO—VB	0.0029096	0.0020978	0.0028958	0.0025040	M	0.58676
VBZ—VBN—DT	0.0016314	0.0016331	0.0014481	0.0014713	M	0.41426
VBZ—VBN—IN	0.0045614	0.0029218	0.0043827	0.0023691	M	0.88898
VBZ—VBN—TO	0.0030156	0.0026660	0.0028980	0.0021945	M	0.80179
WDT—DT—NN	0.0026713	0.0019349	0.0027360	0.0021824	F	0.83900
WDT—MD—VB	0.0014550	0.0014697	0.0014024	0.0013083	M	0.97600
WRB—DT—NN	0.0035177	0.0028049	0.0029347	0.0018175	M	0.39643
WRB—PRP—VBD	0.0013790	0.0019509	0.0016704	0.0023196	F	0.19148

Findings *significant in this study* ($p < 0.05$) are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Table I.6: **All miscellaneous features in present study**

	Gender M		**Gender F**		Gender	Mann-Whitney
Feature	Mean	Std Dev	Mean	Std Dev	Prevalence	p-value
Sentence length (tokens)	26.173	4.237	26.655	3.949	F	0.24840
No. of sentences	147.326	31.152	143.567	26.689	M	0.30530
No. of tokens	3769.787	599.709	3759.260	538.592	M	0.88694
1st person pronoun	0.0010778	0.0009847	0.0010719	0.0010970	M	0.67177
Singular	0.0004910	0.0007667	0.0006099	0.0009267	F	0.56632
Plural	0.0005868	0.0006413	0.0004620	0.0006736	M	0.05839
2nd person pronoun	0.0003059	0.0004212	0.0003511	0.0007255	F	0.50957
3rd person pronoun	0.0131716	0.0093932	0.0133278	0.0087502	F	0.78407
Singular	0.0091763	0.0088748	0.0096805	0.0085672	F	0.44262
Feminine	0.0034325	0.0046678	0.0037717	0.0043508	F	0.25947
Masculine	0.0057439	0.0067667	0.0059088	0.0071111	F	0.64901
Plural	0.0039952	0.0028064	0.0036473	0.0033847	M	0.04151
Contractions	0.0072116	0.0038145	0.0077730	0.0041377	F	0.32913

Findings *significant in this study ($p < 0.05$)* are shaded. See Appendix I.1 at page 287 for detailed explanations of columns.

Appendix J

Findings from machine learning trials

This chapter reports findings of the machine learning algorithm (MLA) trials used for the empirical study in this dissertation. The data collection and preparation efforts described in Chapter 4 resulted in a corpus of the memoranda or legal briefs submitted by 193 students at the end of their first year of law school. This corpus was abstracted into 986 lexical and quasi-syntactic features or attributes, basically relative frequencies of certain function words, parts of speech, and patterns of parts of speech (bigrams and trigrams), for each of 193 instances. (See Section 3.3.1 for a fuller discussion of the operation of MLAs.)

The purpose of these machine learning trials is to see whether MLAs can classify these papers according to authors' self-identified genders using the features described above. The result of the application of each successful MLA should be a concept description. Ultimately, for the concept descriptions to be useful for the broader purpose of this dissertation, they will need to tell us something about how the MLAs succeeded in classifying texts; in other words, they need to be intelligible.

The MLAs applied in this study are those implemented in the WEKA machine learning framework (Hall et al., 2009; "Weka 3: Data Mining Software in Java," n.d.; Witten et al., 2011). For each set of trials, this chapter describes the findings. For each MLA, this includes a brief description of its method of functioning; the attributes with which it can function; necessary transformations of the data set described in Appendix G, if any, to implement the MLA; and identification of parameters used with the MLA in these trials. Because the authors of Witten et al. (2011) are principally responsible for the development and maintenance of WEKA, I have relied on their work for a description of the WEKA tools and for guidance on how to implement them.

First among the trial results, Appendix J.1 reports the findings of the trials using the `Winnow` algorithm, as it was the algorithm used in the Argamon/Koppel 02/03 study. I then report the results of trials with other linear learning approaches in Appendix J.2, instance-based classifiers in Appendix J.3, and support vector machines in Appendix J.4. Finally, Appendix J.5 reports the results of trials with the `NaiveBayes` classifier algorithm. The most successful MLAs and the attributes which most aided their success are summarized in Appendix J.6.

J.1 Trials with the `Winnow` algorithm

I first applied the `Winnow` algorithm (and balanced `Winnow`) because it was the algorithm used in the Argamon/Koppel 02/03 study, which inspired this empirical study and other studies discussed in Chapter 3. Application of both versions of `Winnow` required some preprocessing of the data. Classification of the texts in this study with `Winnow` and balanced `Winnow` achieved one statistically significant result on a trial with the corpus, though this result did not meet the threshold for practical significance.

According to Witten et al. (2011, pp. 129-31), `Winnow` and balanced `Winnow` generate linear models. During the training phase, the algorithm assesses what weight should be assigned to each attribute in the data set. The knowledge representation resulting from it is a linear model that represents the weight assigned to each attribute. When the model encounters an instance in the test set, it multiplies each attribute value in the test instance by its corresponding weight in the model. The results for all attributes are summed, and their category is assigned based on whether the sum exceeds the default threshold. In the basic `Winnow` algorithm, only positive weights can be assigned to attributes. Balanced `Winnow` permits the assignment of positive and negative weights, which can be an advantage in some domains, according to Witten et al. (2011, p. 131).

The WEKA `Winnow` implementations function with nominal or binary values, not with numeric values. The attributes in the data set created for this dissertation (described more thoroughly in Chapter 4 and Appendix G) have numeric values; consequently, they must be transformed so that the attributes are nominal. WEKA provides the means to *discretize* the numerical values in the dataset using a *filter*. Values in each numeric attribute are grouped into *bins*. For example, assume that values for a particular numeric attribute are between 0 and 1 (which is true in the case of this study, though a value of 1 is only theoretically possible): "0.0008399853", "0.00084023", and "0.005340293". The values for that attribute would be grouped together into ranges or *bins*, and the attribute's values would identify a specific range. Assume that the bins are "> 0.00083 & ≤ 0.001", "> 0.001 & ≤ 0.005", and "> 0.005 & ≤ 0.009". Thus, two of the specific numerical values—"0.0008399853" and "0.00084023"—would be binned together, and their values would be transformed to the value "> 0.00083 & ≤ 0.001". No instance has a value greater than 0.001 and less than or equal to 0.005, and so no instances would receive that value. One value, "0.005340293", would be transformed to the value "> 0.005 & ≤ 0.009".

Discretizing numeric values into nominal bins requires decisions about how many bins there should be and how their boundaries will be determined. For use with the `Winnow` algorithm, I performed unsupervised binning using both the equal-frequency and equal-width binning approaches. Equal-width binning creates bins of identical numerical widths; as a result it may create bins with very many or very few members. Equal-frequency binning, on the other hand, spreads the instances so that an equal number of them falls into each bin; the researcher specifies the number of bins. In the latter case, I tried having WEKA use 10 bins and 5 bins and evaluated both. In the former case, I allowed WEKA to determine the optimal number of bins to use.

Witten et al. (2011) did not identify any other pitfalls for the use of either version of `Winnow`. I ran both `Winnow` and balanced `Winnow` with the following parameters: $\alpha = 2.0$, $\beta = 0.5$, Θ equal to the number of attributes or 986, default weight assigned to each attribute at the beginning of the training = 2.0, number of iterations = 5. The results for the full data set, with all 986 attributes, appear in Table J.1 in the section labeled "Full 986-feature set." Results are presented as observed agreement, the percentage of instances classified correctly. The first column of results shows the performance of the `ZeroR` classifier, which is the most frequent class baseline and thus always yields

53.87%. The second column shows the performance of `Winnow` and the third of balanced `Winnow`. I also prepared a smaller set of attributes using the `ClassifierSubsetEval`, as described in Section 4.4.2, and applied the algorithms with the same settings. Those results appear in Table J.1 in the section labeled "Reduced feature set."

Table J.1: **`Winnow` and Balanced `Winnow` Performance**

Data set	Baseline (ZeroR)	Winnow	Balanced Winnow
Full 986-feature set			
Equal Frequency Binning (10 bins)	53.87	49.48	50.06
Equal Frequency Binning (5 bins)	53.87	54.28	52.89
Equal Width Binning	53.87	50.24	46.63 •
Reduced feature set			
Equal Frequency Binning (10 bins)	53.87	56.93	56.60
Equal Frequency Binning (5 bins)	53.87	60.32	61.85 ∘
Equal Width Binning	53.87	57.39	59.24

Improvement (∘) or degradation (•) from baseline, statistically significant at $p<0.05$

As Table J.1 shows, `Winnow` produced results statistically better than the baseline `ZeroR`, but only using the reduced feature set. None of these results reached the level of practical significance (66.00%) described in Section 4.4.2. The approach to discretization binning appeared to have small effects.

`Winnow` is only one of the machine learning algorithms in WEKA that generates linear models. The next section presents the results of three others.

J.2 Trials with other linear approaches

As I noted above, `Winnow` generates a linear model, meaning that it generates a knowledge representation in the form of a linear regression equation. WEKA implements several other MLAs that produce linear models, and I tried my data against three of them: `Logistic`, `SimpleLogistic`, and `VotedPerceptron`. Each of these algorithms achieved results better than `Winnow`, two statistically and one practically significant.

The `Logistic` and `SimpleLogistic` algorithms build models via logistic regression (Witten et al., 2011, p. 126, 467) that estimate class probabilities. As its name suggests, `SimpleLogistic` builds a logistic model. `Logistic` is designed to address the problem of *overfitting*, which I noted in the discussion on page 130 is commonly of concern in machine learning:

> In [overfitting], the learning algorithm adapts so well to the given data, that noise or particularities of the specific sample are also encoded by the learned model. It results in reduced performance when the task is the generalization to unseen data, as well as producing an overly complex model which may consume unnecessary learning time and computational resources (Wu & Shapiro, 2006, p. 433).

In short, overfitting results in a learned model that is less useful outside the context of the samples in the current study. Proper use of cross-validation helps to address this concern. A learning algorithm can also be designed to address the problem of overfitting; and in WEKA, `Logistic` is designed to do so by the means described in Witten et al. (2011, p. 467).

Both `Logistic` or `SimpleLogistic` can learn with numeric attributes (like the data prepared for this study), so no transformations of the data were necessary before application of these algorithms. Witten et al. (2011) did not identify any other pitfalls for the use of these MLAs. I ran `SimpleLogistic` with the following parameters: No error on probabilities; heuristic stop = 50; maximum number of boosting iterations = 500; no fixed number of boosting iterations; no AIC; cross-validation = true; no weight trimming. I ran `Logistic` with the following parameters: no maximum number of iterations (-1); ridge value = 1.0E-8.

Table J.2: **Other Linear Model Performance**

Data set	Baseline (ZeroR)	Logistic	Simple Logistic	Perceptron
Full 986-feature set	53.87	58.81	52.56	53.76
Reduced feature set	53.87	55.79	66.76 ○	65.29 ○

Improvement (○) or degradation (●) from baseline, statistically significant at $p<0.05$

The perceptron algorithm, described by Witten et al. (2011, p. 129) as the "grandfather of neural networks," produces a model that describes a "hyperplane" that separates the instances into the two classes rather than generating probability estimates like `SimpleLogistic` and `Logistic`. The `VotedPerceptron` implementation in WEKA is designed to address a problem with the perceptron algorithm: that the model it learns is very much subject to the order of the instances to which it is exposed (Witten et al., 2011, p. 232). `VotedPerceptron` accepts numeric attribute data and Witten et al. (2011) did not identify any other pitfalls for its use. I ran it with the following parameters on the entire 986-attribute data set: exponent = 1.0; maximum alterations = 10,000; number of iterations = 1; random number seed = 1.

The results of applying all three algorithms to the 986-feature full data set appear in in Table J.2 in the row labeled "Full 986-feature set." I also prepared a smaller set of attributes using the `ClassifierSubsetEval`, as described in Section 4.4.2. Those results appear in Table J.2 in the row labeled "Reduced feature set."

No algorithm achieved statistically better results than baseline did using the entire 986-feature data set. Using the reduced feature set, however, `SimpleLogistic` and `Perceptron` achieved results that were statistically significant. `SimpleLogistic` performed better than either other algorithm, achieving practical significance, with its best performance at 66.76%.

The concept description in the linear models described in this section and Appendix J.1 is an equation generated by an analysis of the training set. Once the equation is created, there is no further need for the training instances. At classification time, the attributes of the test instance are multiplied by the coefficients in the equation, and the instance is placed in one class or the other depending on whether the result meets a certain threshold. At the other end of the spectrum are instance-based algorithms, which do not build "models" at all.

J.3 Trials with instance-based classifiers

Instance-based classifiers are different than those previously discussed, because they do not build a model (linear or otherwise) to represent the class boundaries (Witten et al., 2011, p. 131-38). This section reports the results of trials with two instance-based learning algorithms: `IB1` and `IBk`. The latter achieved performance that was better than the most frequent class baseline; though that improvement was statistically significant, it was not meet the standard of practical significance discussed above.

In the simplest instantiation of instance-based learning, each test instance is compared to each member of the training set at classification time: the mathematical distance between them is calculated and the test instance is given the class of the training instance that is "closest" to it in mathematical terms, called its "nearest neighbor" (Witten et al., 2011, p. 131-38). In this type of instance-based learning, the concept description is the whole training set. This approach has the virtue that new training instances can be added easily: In the linear model environment, a new training instance necessitates generation of a new equation to function as the concept description. In the instance-based environment, a new training instance is simply one more instance against which the test instance is compared.

Instance-based algorithms suffer from a variety of problems, including computational complexity, vulnerability to "noisy" data, and the lack of an interpretable concept description. First, instance-based machine learning algorithms are sometimes described as "lazy," because they do all their work at the time of classification. They might better be described as "procrastinators," because the work they do at the time of classification requires heavier lifting than do the linear models in the algorithms discussed above. Support vector machines, described in Appendix J.4, go some way so solving that problem. Computational complexity has not posed a problem in this study because of the (relatively) small number of attributes and instances at stake here. Another possible problem is "noisy" data. A single training instance with erroneous values, or a small number of them, can disrupt effective classification of new instances. One solution to that problem is to use a small number —k—of neighbors (e.g., five) instead of the single nearest neighbor and assign the class based on how the majority of these k nearest neighbors are classified. The trials described in this section include both the `IB1` basic nearest neighbor classifier algorithm and the `IBk` k-nearest-neighbor classifier.

From the perspective of this dissertation, a more significant problem with instance-based classifiers is the lack of a concept description outside of the actual classification task. When the model for nearest-neighbor learning is the training set, and we can't know how an instance will be classified until the time of classification, the learning algorithm starts to look like a "black box"—a process the inputs and outputs of which we can examine, but the operation of which will always be a mystery. Thus, if the results of an instance-based classifier show a statistically significant improvement over the most frequent class baseline, I will not be able to describe what attributes of the texts figured prominently in their classification. I hope to use the results of the trials that use reduced feature sets to identify features that can be used in a theoretical explanation of the results.

Witten et al. (2011, p. 132) note if the numeric values of the training attributes are at different scales, small relative effects in one attribute can dwarf large relative effects in another. As a solution, they recommend that the attribute values be *normalized*, that is, scaled so that their values all lie between 0 and 1 (p. 437). For these instance-based learning trials, I normalized the attribute values in my data using the `Normalize` filter in WEKA with default settings. Witten et al. (2011) did not identify any other pitfalls with these algorithms, so I employed each with its default settings: `IB1` does not have any user-selected options; for `IBk`, I set k to 5 neighbors, used cross validation, did not use distance weighting, mean squared error, or a window size paremeter, and applied the `LinearNNSearch` algorithm for identifying nearest neighbors. I ran these MLAs against the entire, 986-attribute data set. The results of these trials appear in Table J.3 in the row labeled "Full 986-feature set."

Table J.3: **Instance-based classifier performance**

Dataset	Baseline (ZeroR)	IB1 (nearest neighbor)	IBk (k-nearest)
Full 986-feature set	53.87	48.04	50.19
Reduced feature set	53.87	55.91	62.93 ○

Improvement (○) or degradation (●) from baseline, statistically significant at $p<0.05$

I used the `ClassifierSubsetEval` feature selector to prepare a smaller set of attributes for each algorithm, as described in Section 4.4.2. Those results appear in Table J.3 in the row labeled "Reduced feature set."

Only the `IBk` trial that used the reduced data set produced observed agreement that was significantly better than the baseline. Its best performance was 62.93% observed agreement, but it never achieved practical significance by the standards set out in Section 4.4.2.

The designers of machine learning algorithms are pragmatic, not dogmatic, in their methods. Consequently, there is a class of MLAs intended to take advantage of some of the strengths of both the linear and the instance-based methods.

J.4 Trials with support vector machines

Support vector machines (SVMs) combine aspects of instance-based and linear models. At the time of training, the SVM generates a special kind of linear model, a hyperplane that separates the classes by the widest margin (Witten et al., 2011, p. 223-227). The margin is calculated using the nearest-neighbor concept from instance-based learning. At classification time, the linear model is applied to the test instance. This section reports the results of trials with two SVM algorithms: `SMO`, which delivered one of the top four trials in this entire study; and `SPegasos`, which had no practically significant result, though it did have one result statistically better than baseline.

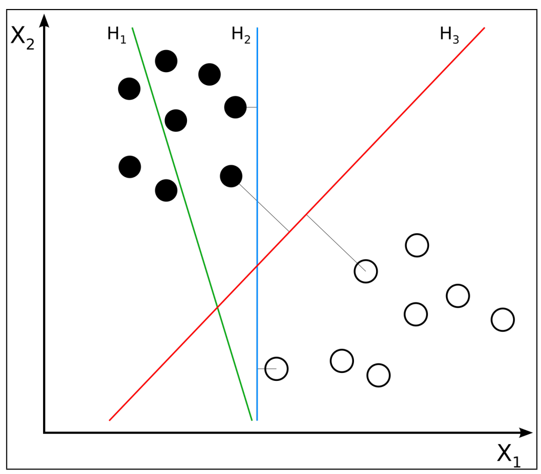


Figure J.1: **Illustration of a hyperplane separating instances with two attributes.** Source: ZachWeinberg (n.d.).

Figure J.1 illustrates the concept of the *maximum-margin hyperplane*. Here, the instances are represented by circles, filled in the case of one class and open in the case of the other. The hyperplane labeled "H_1" does not discriminate the instances into the classes: some of the filled circles are on the same side of it as the open circles. It is therefore not a good candidate for a classifier model. The hyperplane labeled "H_2" does discriminate the instances into the proper classes and could represent a linear model for a classifier as described in Appendix J.1 or Appendix J.2 above. However, it may not be as effective as we would like: a filled-circle instance might appear just to the right of the right-most existing filled instance, and therefore on the wrong side of H_2. An SVM would describe the hyperplane labeled "H_3," which separates the instances into the two classes while itself maintaining the greatest mathematical distance from the instances in each class. The position of H_3 is calculated by maximizing its mathematical distance from a small number of instances nearest to either side of it, in this case, the filled and open circles connected to it with thin lines.

In this way, the SVM is like the instance-based models above, except that the calculation of the distance happens during training. At classification time, the linear model is applied to the test instance just as in the linear algorithms above, much reducing the computational complexity of the classification task.

I used two SVMs from the WEKA tool-suite, `SMO` and `SPegasos`. These algorithms automatically *normalize* attribute values, scaling them so that they all lie between 0 and 1 (Witten et al., 2011, p. 437). With `SMO`, I used default parameters: no logistic models; complexity = 1.0; checks on; ε = 1.0E-12; normalized training data; kernel = `Polykernel` (-C 250007 -E 1.0); folds = -1; tolerance = 0.0010. I used default parameters for `SPegasos` as well: normalization and replaced values on; 500 epochs; λ = 0.0001; hinge loss function. I ran these MLAs on the entire 986-feature data set, and the results appear in Table J.4 in the row labeled "Full 986-feature set." I also prepared a smaller set of attributes using the `ClassifierSubsetEval`, as described in Section 4.4.2. Those results appear in Table J.4 in the row labeled "Reduced feature set."

Table J.4: **Support vector machine performance**

Dataset	Baseline (ZeroR)	SMO	SPegasos	
Full 986-feature set	53.87	59.49	57.91	
Reduced feature set	53.87	71.57 ○	62.12	○

Improvement (○) or degradation (●) from baseline, statistically significant at $p<0.05$

As Table J.4 shows, both of the SVM algorithms resulted in statistically significant improvements over the most frequent class baseline provided by `ZeroR`. In fact, `SMO` delivered the second-highest observed agreement—71.57%—of any algorithm in this dissertation. `SPegasos` was less impressive, delivering one result that was barely of statistical significance at 62.12% observed agreement.

Until now, support vector machines and linear methods have performed in something of a dead heat in these trials, leaving `Winnow` in the dust. And Table 5.7 gives a sense of how these two categories of algorithm produced two of the four practically significant results in this dissertation. However, the same table shows that a single algorithm holds the top spot: `NaiveBayes`.

J.5 Trials with Naive Bayes models

Machine learning algorithms (MLAs) in the Naive Bayes category take advantage of Bayes' theorem to generate a concept description in the form of a *Bayesian belief network*. In mathematical terms, a Bayesian belief network is a "directed, acyclic graph in which nodes represent domain variables which have a finite number of possible values" (Zheng & Tang, 2005, p. 282). The "graph" mentioned here is a mathematical concept and should not be confused with a visual chart (though graphs are sometimes represented graphically). The result of applying a Bayesian classifier to training data is a set of conditional probabilities associated with classes given the priors and probabilities of the priors given the class labels. The model consists of a set of variables, arcs between them, and probabilities on each of the arcs.

Naive Bayes classifiers take advantage of a simplifying assumption, that the attributes are not statistically related to each other. This makes the classifiers computationally efficient, but it leaves them vulnerable to error when the attributes selected are redundant or closely related to each other. One means of dealing with this is to pre-process the attributes, selecting those that are statistically unrelated to each other and that yield the best performance. A second concern is that the standard Naive Bayes probabilistic classifier implemented in WEKA (`NaiveBayes`) anticipates that numerical attributes will have a *Gaussian* distribution.[1] When values are plotted on a chart, a Gaussian distribution takes the familiar form of the bell curve, a symmetrical shape with a relatively high center and tails that trail off to the right and left. Though some of the

[1] This is commonly called a *normal* distribution, but I feel that term suffers from a number of problematic associations.

attributes in the corpus here have distributions that are Gaussian or nearly Gaussian, the great majority do not. The `NaiveBayes` can compensate for distributions that are non-Gaussian using a tool called a *kernel estimator*.[2]

I ran `NaiveBayes` both with and without the kernel estimator on the full data set of 986 attributes. The results appear in Table J.5 in the row labeled "Full 986 feature set." I also prepared a smaller set of attributes for each version of Naive Bayes using the `ClassifierSubsetEval`, as described in Section 4.4.2. Those results appear in Table J.4 in the row labeled "Reduced features 'ClassifierSubet."' Finally, I did a second round of attribute selection with the `WrapperSubsetEval` attribute evaluator (parameters: NaiveBayes -F 5 -T 0.01 -R 1 –) and the `SubsetSizeForwardSelection` search method (parameters: -I -K 50 -T 0 -F 5 -S 1 -Z false -E). Witten et al. (2011, p. 314) specifically recommended this attribute selection for Naive Bayes. The attribute search algorithm uses 10-fold cross validation. I identified the attributes identified as useful in at least one of the folds. I created only one such set of attributes for the corpus, and used it with `NaiveBayes` with and without the kernel estimator. The results appear in Table J.5 in the row labeled "Reduced features 'WrapperSubset."'

Table J.5: `NaiveBayes` **classifier performance**

Dataset	Baseline (ZeroR)	NBayes	NBays with Kernel Est.	
Full 986-feature set	53.87	50.61	51.84	
Reduced attributes 'ClassifierSubset'	53.87	63.74 ○	49.61	
Reduced attributes 'WrapperSubset'	53.87	73.19 ○	70.42	○

Improvement (○) or degradation (●) from baseline, statistically significant at $p<0.05$

`NaiveBayes` using the reduced feature set selected with `WrapperSubset` was the best performer among all the algorithms discussed in this appendix. It delivered the highest observed agreement: 73.19%. One of the results with the reduced feature set selected with `ClassifierSubset` was also statistically, though not practically, significant. This algorithm performed better consistently without the kernel estimator discussed above.

[2] The kernel estimator is an alternative to basic `NaiveBays`, advised by Witten et al. (2011) for feature sets where continuous variables exhibit a non-parametric distribution. See the discussion in Section 4.4.1 regarding the lack of Gaussian distribution for most features in this study.

Given the high performance of `NaiveBayes` with the reduced feature set produced by `WrapperSubset`, the features in that data set will likely be of interest in interpreting these results. See the discussion of them in Section 5.3.3.

J.6 Summary of machine learning trial results

See Table 5.7 for a summary of the machine-learning trials that provided significant improvement over baseline.

See Table 5.8 for the reduced feature sets used by `SimpleLogistic`, `SMO`, and `NaiveBayes` to achieve statistically and practically significant results.